RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING

A Thesis submitted to Gujarat Technological University

for the Award of

Doctor of Philosophy

in

Computer Science

By

Monica Gupta
Enrollment No: 199999902503

under supervision of

Dr. Sohil Pandya

GUJARAT TECHNOLOGICAL UNIVERSITY

AHMEDABAD

[July – 2026]

DECLARATION

I declare that the thesis entitled **RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING** submitted by me for the degree of Doctor of Philosophy is the record of research work carried out by me during the period from December 2020 to July 2026 under the supervision of Dr. Sohil Pandya and this has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or other institution of higher learning.

I further declare that the material obtained from other sources has been duly acknowledged in the thesis. I shall be solely responsible for any plagiarism or other irregularities, if noticed in the thesis.

Signature of the Research Scholar:  Date: 23-July-2026

Name of Research Scholar: **Monica Gupta**

Place: **Ahmedabad**

CERTIFICATE

I certify that the work incorporated in the thesis **RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING** submitted by **Monica Gupta** was carried out by the candidate under my supervision/guidance. To the best of my knowledge: (i) the candidate has not submitted the same research work to any other institution for any degree/diploma, Associateship, Fellowship or other similar titles (ii) the thesis submitted is a record of original research work done by the Research Scholar during the period of study under my supervision, and (iii) the thesis represents independent research work on the part of the Research Scholar.

Signature of the Research Supervisor:  Date: 23-July-2026

Name of Supervisor: **Dr. Sohil Pandya**

Place: **Anand**

Course-work Completion Certificate

This is to certify that **Monica Gupta** enrolment no. **199999902503** is enrolled for PhD program in the faculty **Computer Science** of Gujarat Technological University, Ahmedabad.

(Please tick the relevant option(s))

☐ ~~He/She has been exempted from the course-work (successfully completed during M.Phil Course)~~

☐ ~~He/She has been exempted from Research Methodology Course only (successfully completedduring M.Phil Course)~~

✓ She has successfully completed the PhD course work for the partial requirement for the award of PhD Degree. His/ Her performance in the course work is as follows-

Grade Obtained in Research Methodology [PHD20-01]	Grade Obtained in Research and Publication Ethics [PHD20-02]	Grade Obtained in Self-Study Course/ Contact Program [PHD20-03]
BB	AA	AB

Supervisor's Sign

Dr. Sohil Pandya

Originality Report Certificate

It is certified that PhD Thesis titled **RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING** by **Monica Gupta** has been examined by us. We undertake the following:

a. Thesis has significant new work / knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless itis placed under quotation marks and duly referenced.

b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.

c. There is no fabrication of data or results which have been compiled / analyzed.

d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

e. The thesis has been checked using **Turnitin (**copy of originality report attached) and found within limits as per GTU Plagiarism Policy and instructions issued from time to time (i.e. permitted similarity index <=10%).

Signature of the Research Scholar:  Date: 23-July-2026

Name of the Research Scholar: **Monica Gupta**

Signature of the Research Supervisor:  Date: 23-July-2026

Name of the Research Supervisor: **Dr. Sohil Pandya**

Place: **Anand**

Monica Gupta

RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING

RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING

PHD Thesis Plagiarism Report

Gujarat Technological University

Document Details

Submission ID
trn:oid:::1:3299102981

Submission Date
Jul 20, 2025, 10:43 PM GMT+5:30

Download Date
Jul 20, 2025, 10:49 PM GMT+5:30

File Name
PREDICTION_IN_THE_CAMPUS_PLACEMENT_THROUGH_MACHINE_LEARNING.docx

File Size
7.7 MB

110 Pages

24,611 Words

146,275 Characters

turnitin Page 1 of 122 - Cover Page Submission ID trn:oid:::1:3299102981

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- 214 Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- 1 Missing Quotations 0%
  Matches that are still very similar to source material
- 1 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
6% Publications
4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Ph.D. Thesis Non-Exclusive License to

GUJARAT TECHNOLOGICAL UNIVERSITY

In consideration of being a Research Scholar at Gujarat Technological University, and in the interests of the facilitation of research at the University and elsewhere, I, **Monica Gupta** having **19999992503** hereby grant a non- exclusive, royalty free and perpetual license to the University on the following terms:

a. The University is permitted to archive, reproduce and distribute my thesis, in whole or in part, and/or my abstract, in whole or in part (referred to collectively as the "Work") anywhere in the world, for non-commercial purposes, in all forms of media;
b. The University is permitted to authorize, sub-lease, sub-contract or procure any of the acts mentioned in paragraph (a);
c. The University is authorized to submit the Work at any National / International Library, under the authority of their "Thesis Non-Exclusive License";
d. The Universal Copyright Notice (©) shall appear on all copies made under the authority of this license;
e. I undertake to submit my thesis, through my University, to any Library and Archives. Any abstract submitted with the thesis will be considered to form part of the thesis.
f. I represent that my thesis is my original work, does not infringe any rights of others, including privacy rights, and that I have the right to make the grant conferred by this non-exclusive license.
g. If third party copyrighted material was included in my thesis for which, under the terms of the Copyright Act, written permission from the copyright owners is required, I have obtained such permission from the copyright owners to do the acts mentioned in paragraph (a) above for the full term of copyright protection.
h. I understand that the responsibility for the matter as mentioned in the paragraph (g) rests with the authors / me solely. In no case shall GTU have any liability for any acts/ omissions / errors / copyright infringement from the publication of the said thesis or otherwise.
i. I retain copyright ownership and moral rights in my thesis, and may deal with the copyright in my thesis, in any way consistent with rights granted by me to my University in this non-exclusive license.
j. GTU logo shall not be used /printed in the book (in any manner whatsoever) being published or any promotional or marketing materials or any such similar documents.
k. The following statement shall be included appropriately and displayed prominently in the

book or any material being published anywhere: “The content of the published work is part of the thesis submitted in partial fulfilment for the award of the degree of Ph.D. in **Computer Science** of the Gujarat Technological University”.

l. I further promise to inform any person to whom I may hereafter assign or license my copyright in my thesis of the rights granted by me to my University in this nonexclusive license. I shall keep GTU indemnified from any and all claims from the Publisher(s) or any third parties at all times resulting or arising from the publishing or use or intended use of the book / such similar document or its contents.

m. I am aware of and agree to accept the conditions and regulations of Ph.D. including all policy matters related to authorship and plagiarism.

Date: 23-July-2026

Place: **Ahmedabad** Signature of the Research Scholar

Recommendation of the Research Supervisor: **Recommended**

Signature of the Research Supervisor

Thesis Approval Form

The viva-voce of the PhD Thesis submitted by Monica Gupta (Enrollment No. : 199999902503) entitled **RECRUITMENT AND SELECTION PREDICTION IN THE CAMPUS PLACEMENT THROUGH MACHINE LEARNING** was conducted on **23-July-2026, Thursday** at Gujarat Technological University.

(Please tick any one of the following option)

☑ The performance of the candidate was satisfactory. We recommend that he/she be awarded the PhD degree.

☐ Any further modifications in research work recommended by the panel after 3 months from the date of first viva-voce upon request of the Supervisor or request of Independent Research Scholar after which viva-voce can be re-conducted by the same panel again.

(briefly specify the modifications suggested by the panel)

☐ The performance of the candidate was unsatisfactory. We recommend that he/she should not be awarded the PhD degree.

(The panel must give justifications for rejecting the research work)

Dr. Sohil Pandya
Name and Signature of Supervisor with Seal

Dr. Neeraj Bhargava
1) (External Examiner 1) Name and Signature

Dr. Jigisha Patel
2) (External Examiner 2) Name and Signature

ABSTRACT

Campus placements serve as a vital link between higher education outcomes and organisational talent pipelines in a labour market that is becoming more and more competitive. Conventional hiring methods on college campuses, however, frequently ignore more comprehensive measures of employability like project portfolios, extracurricular leadership, soft-skill competency, and recruiter feedback in favour of static academic metrics and aptitude test results. In order to anticipate and eventually enhance student placement performance, this thesis proposes and empirically verifies a data-driven framework based on supervised machine learning algorithms.

This study investigates the application of ML techniques in recruitment and selection during campus placements. To forecast student employability, it assesses data from over 60 IT organizations using various criteria, including project experience, technical proficiency, academic accomplishment, certifications, and soft skills. Candidates are evaluated and classified based on their placement eligibility using a number of supervised and probabilistic machine learning models.

The study compares the techniques like Random Forest, Support Vector Machine, Logistic Regression, Gradient Boosting, Decision Tree, and k-Nearest Neighbor models using a data set collected from some recruiters on a confidential basis, as well as some data from Kaggle. The data was then generated using the appropriate algorithm while keeping the 14 parameters in mind. Ensemble and kernel-based approaches achieve headline accuracies above 99 percent.

The study exhibits operational integrations beyond predictive performance, including a recommendation module that recommends targeted up-skilling pathways to students with sub-threshold probabilities, an adaptive assessment engine that continuously improves test blueprints, and a role-specific scoring dashboard for recruiters. According to scenario simulations, implementing the system could reduce recruiter screening time by almost half and increase overall placement coverage. Throughout, ethical measures are incorporated, such as opt-in transparency interfaces, anonymisation pipelines, and fairness audits across

gender and socioeconomic levels.

The study emphasizes the significance of analytical based decision-making in talent acquisition while also drawing attention to the inadequacies of traditional hiring processes. Recruiters may enhance recruiting results by employing machine learning to conduct more complete, unbiased, and objective evaluations. In addition to assisting with campus recruiting, this study creates a foundation for future research in deep learning and adaptive models, which will eventually match academic education to changing industrial demands.

The thesis concludes with strategic recommendations for universities and industry partners and proposes a road to fair, successful, and evidence-based college recruiting ecosystems by combining machine learning rigour with human-centered design.

Acknowledgement

With immense gratitude and humility, I acknowledge the support, guidance, and encouragement that I have received during the course of my doctoral research.

I am profoundly thankful to my research supervisor, **Dr. Sohil Pandya**, for his unwavering guidance, thoughtful insights, and continuous encouragement throughout this journey. His expertise and patient mentorship were crucial in shaping the direction and success of this research work.

I would like to express my heartfelt gratitude to my Doctoral Progress Committee, **Dr. Dhiren B. Patel** and **Dr. C. K. Kumbharana**, for their valuable insights, constructive feedback, and continuous support throughout the course of my research. Their thoughtful comments and critical reviews not only enhanced the quality of my work but also guided me to refine my approach and deepen my understanding for this research.

I extend my sincere appreciation to **Gujarat Technological University**, Ahmedabad, for providing the academic platform and resources necessary for carrying out this research.

I also wish to express my deep gratitude to my family and friends for their unwavering encouragement and emotional support throughout this journey. A heartfelt thanks to my husband, **Mayank**, whose patience, understanding, and constant motivation helped me persevere, especially during the most demanding phases. I am equally grateful to my daughter, **Shivanya**, for her patience and silent support, which meant a great deal during this period.

A special word of appreciation goes to **Mr. Rutvik Panchal**, for his consistent assistance and readiness to help whenever I faced challenges in my research. His contribution and support have been truly valuable.

Finally, I dedicate this work to all those who believe in the power of research and innovation to improve systems and impact lives positively.

Monica Gupta

Table of Contents

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
ATS	Applicant Tracking System
AUC	Area Under the Curve
CBL	Cascade Bi-Level (feature selection)
CV	Cross-Validation
DEI	Diversity, Equity & Inclusion
EDM	Educational Data Mining
F1	F-Measure
FN	False Negative
FP	False Positive
GPU	Grade Point Average
KNN	k-Nearest Neighbors
ML	Machine Learning
MOOC	Massive Open Online Course
NLP	Natural Language Processing
PSO	Particle Swarm Optimisation
ROC	Receiver-Operating-Characteristic
SVM	Support Vector Machine
TPR	True Positive Rate (Sensitivity / Recall)
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

List of Figures

List of Tables

CHAPTER – 1

INTRODUCTION

1.1 Overview

A company's employees have the power to make or break it. They are in charge of carrying out your business's operations, influencing clients, and generating income. Herein lies the significance of hiring and choosing individuals in HRM (Human Resource Management), which determines the demands of our business and matches them appropriately. Companies may suffer from a poor recruiting and selection process in a number of ways. Employing the incorrect personnel may have an impact on many facets of your everyday business operations, in addition to lowering earnings due to inefficiency.

A study assesses the elements that affect hiring and selection procedures for various companies looking to expand. Organisational growth is greatly impacted by hiring practices and staff performance, according to content analysis. The study underlines the prominence of appropriate recruitment practices in selecting the most qualified candidate. The growth of an organization depends on having knowledgeable and qualified personnel. The study also emphasizes the significance of having a pleasant workplace for enhancing employee performance and productivity. It also highlights how important employee incentive is for increasing retention and reducing turnover. Thus, the growth of every business depends on selecting and employing the right personnel. [1] (Abbasi, S. G. *et.al*, 2022)

Recruitment and selection are two of the most crucial facets of human resources management. It includes identifying, attracting, and selecting the best candidate for a position. Selection and recruitment may be costly and time-consuming procedures. However, it is a crucial component of an organization.

- **Finding the Correct Talent :**

Finding and hiring the best personnel is one of the most important parts of the process. Any organization's personnel are essential to its success. The organisation may advance by

bringing in fresh perspectives, knowledge, and abilities from the proper personnel. Employing the right people may help businesses boost output and profitability. To attract the right candidates, organizations must be open and honest about their job requirements, including the qualifications and abilities needed for the role. They must also have a good recruitment strategy. This may include posting job openings on employment boards, social media sites, and other pertinent outlets. Through this approach, companies may potentially draw in a substantial number of applicants who possess the necessary training and expertise.

- **Making Certain a Diverse Staff :**

In the contemporary, globalised world, variety and addition are critical to any establishment's success. New viewpoints, concepts, and experiences can aid organisations in problem-solving and innovation. However, it might be difficult to create a diverse workforce. A diverse workforce may be ensured by organisations through recruitment and selection. A fair and open hiring procedure can help companies draw applicants from a variety of cultural and ethnic backgrounds. Any prejudices that could be present in the hiring process can also be found and removed by them. By doing this, organisations may ensure that all applicants, irrespective of their gender, race, ethnicity, or other personal attributes, are on an even playing field.

- **Lowering the rate of staff turnover :**

For businesses, it is essential since it might result in lost knowledge, productivity, and hiring expenses. Choosing the proper individuals via recruitment and selection may help businesses save money on hiring expenses, boost productivity, and enhance job satisfaction and retention over the long run.

- **Increasing the Culture of the Organization :**

For an organisation to succeed, its culture—which is characterized by common values, beliefs, and practices—is essential. Productivity, motivation, and employee engagement are all increased by a healthy workplace culture. Organisations may cultivate a culture that values creativity, teamwork, and diversity by using recruitment and selection to assist them draw in the best people. Through the identification of people who share these values, organisations may establish a culture that is in line with their goals and objectives [1].

Globalization and the growth of the IT industry have changed the nature of success in higher education. Institutions are now ranked according to average wages and successful

job placements, emphasizing the significance of campus placement and institutional accountability for successful students. The new model and approach for students to acquire demanding employment offers from respectable worldwide firms is discussed in this research, with a focus on the intellect and hard work needed for progress and sustainability within the company.[2] **(Lena, R., 2024)**

1.2 Traditional Campus Recruitment Practices

The standard approaches and procedures used to find and hire workers are referred to as traditional recruiting methods. These techniques, which have existed for millennia, consist of:

- **Print advertising** is the practice of advertising job opportunities in newspapers or on job boards in order to attract a large number of possible applicants.
- **Internal Hiring:** Using recommendations from current staff members to fill vacancies.
- **Local Employment Offices:** To reach a larger audience, post job vacancies at local employment offices.
- **Temp Agencies:** Shortlisting and identifying qualified applicants through temporary job agencies.
- **Headhunting** is the process of finding applicants via personal networks and relationships.
- **Organizing In-Person Interviews:** Asking applicants to come to the company's offices or a neutral setting to speak with a hiring manager or a panel of interviewers face-to-face.
- **Examining Resumes and Cover Letters**: Examining applicants' resumes and cover letters to determine their suitability for the role and level of qualifications.
- **Conducting Tests or Assessments:** Conducting aptitude tests or other evaluations to gauge the competencies of applicants.
- **Verifying References:** To learn more about a candidate's work history and performance, get in touch with their references, such as previous managers or co-workers.
- **Referral Initiatives:** Referral programs encourage existing workers to recommend friends or co-workers for available opportunities since these referrals are frequently more successful and dependable recruits.

1.2.1 Advantages of Traditional Method

Benefits of traditional hiring practices include familiarity, a personal connection, and established procedures. They enable businesses to use in-person interviews to evaluate candidate's personalities and organisational fit. However, their restricted reach, time-consuming nature (due to in-person interviews), possibility for unconscious bias, and lack of data may make it challenging to make well-informed recruiting judgements. However, more precise and thorough data is provided by modern approaches, which makes them more appropriate for a wide range of candidate pools.

1.3 Contemporary Hiring Practices

Modern recruiting tactics are the latest approaches and plans for identifying and employing the most qualified applicants for a position. To increase productivity, improve the applicant experience, and streamline the hiring process, these strategies make use of technology and creative thinking. Here are some important contemporary hiring practices

- **AI Technology :**

AI streamlines the employment process by automating procedures like candidate matching, screening, and filtering. It also helps identify the top candidates by looking at their credentials and experience.

- **Social Media Sites:**

To engage with potential workers and build the employer brand, job openings are shared on Social networking sites such as Facebook, LinkedIn, Twitter and many others. This approach can reach a large audience and target particular demographics.

- **Applicant tracking systems (ATS) :**

ATS software facilitates the management and screening of job applications, hence facilitating the identification of the most qualified applicants. Additionally, by automating administrative duties, it saves time and money.

- **Virtual Reality (VR) Communication :**

Candidates can communicate with the firm remotely through virtual interviews made

possible by VR technology. This approach offers a more engaging experience and might be especially helpful for applicants who live far away or are from other countries.

- **Mobile Recruiting :**

Job postings are posted, prospects are contacted, and interviews are conducted using mobile devices. Reaching younger generations, who are more inclined to utilise mobile devices for job searches, is especially successful with this approach.

- **Talent Pool :**

This approach entails keeping track of prior applicants who had promise but were not recruited. These applicants can be quickly contacted and given consideration for the role when a new opening occurs.

- **Boomerang Workers :**

This tactic entails recruiting back former workers who parted ways amicably. Since these workers are already familiar with the corporate culture and are likely to blend in with ease, this approach may work.

1.4 Internal Recruitment

In this strategy, internal promotions are made. It is less expensive and may boost employee engagement and motivation by demonstrating a dedication to their professional development.

- **Collaborative recruiting :**

In this approach, executives and line managers participate in the recruiting procedure. It guarantees that the chosen applicant satisfies the demands and anticipations of all parties involved.

- **Interviews via Video :**

Remote interviews may be conducted by video, which saves time and money. They also provide the applicant a more intimate experience.

- **Employer Branding :**

This strategy entails developing a strong employer brand in order to recruit top people. It entails demonstrating the corporate culture, values, and perks to prospective employees.[2]

1.5 Challenges in Existing Recruitment Methods

There are several challenges in the existing recruitment procedure:

- **Talent Shortages in Key Industries :**

Specialized skill shortages will arise when administrations adopt the latest technologies like artificial intelligence, cyber security or machine learning, particularly in the engineering, technology, and healthcare sectors. According to McKinsey, 87% of companies will have a skills gap by 2025, which might impede innovation and growth.

- **Difficulties associated with remote work:**

While remote work allows for greater freedom, it also brings new obstacles in hiring, managing, and maintaining a remote staff. Hiring across time zones and managing cross-cultural teams may make communication difficult and undermine cohesion. According to Gartner, 81% of employers struggle to successfully manage remote teams.

- **Diversity, Equity, and Inclusion :**

Despite growing awareness, many businesses still struggle to create diverse and inclusive teams. Unconscious prejudice can infect recruiting procedures, and non-inclusive hiring practices may turn off potential employees. According to a LinkedIn poll, 67% of job searchers value diversity when picking employment. In 2025, inclusiveness will be a major distinction for businesses seeking to establish innovative, effective teams.

- **Employer Branding in a Competitive Environment:**

Employer branding will become increasingly important in 2025 as candidates grow pickier about their job hunt. With more competition for top personnel, organisations with poor employer brands may lose out to competitors that provide better possibilities or workplace culture. According to Glassdoor research, 72% of recruiters feel that employer branding has a major influence on hiring success.

- **AI and Automation in Recruitment :**

While automation may assist expedite recruiting procedures, over-reliance on AI can result in a lack of customization, severely harming the applicant experience. According to Deloitte, 41% of organisations now utilize AI in recruiting but are struggling to strike the proper mix between efficiency and customization.

- **Employee Retention & Turnover :**

Retaining personnel will remain a big concern, especially as the competition for talent heats up. According to a Gallup research, 51% of employees are actively seeking for new employment options, mostly because they want greater pay, professional growth, and flexibility

- **Budget constraints :**

Economic instability and budget constraints will make recruiting more difficult, since businesses may struggle to invest in cutting-edge recruitment tools or attract top talent. Budget limits may limit access to important tools and platforms, affecting recruiting efficiency.

- **Long hiring processes :**

Long recruiting processes are a major source of frustration, with 60% of job seekers abandoning lengthy applications. Prolonged recruiting delays not only annoy candidates, but also cause organisations to lose out on top personnel.

- **Poor communication during the hiring process :**

A lack of communication between recruiters, hiring managers, and applicants can cause delays and provide a negative candidate experience. Misalignment among internal stakeholders can lead to misunderstanding, missed opportunities, and higher candidate dropout rates. Worse, when conversations or video interviews are not recorded, essential facts may be forgotten or misremembered, making it difficult to follow up on vital issues or keep everyone on track.

- **Sustainability and Social Responsibility :**

More applicants, particularly younger generations, are assessing potential employers'

commitment to sustainability and corporate social responsibility (CSR). According to a Deloitte poll from 2024, 61% of millennials evaluate a company's environmental and social goals when choosing a job.[3]

1.6 Emergence in Recruitment Prediction through Machine Learning

The combination of technology and data has resulted in previously unheard-of levels of precision, efficiency, and strategic insight, transforming the recruiting environment. In order to enhance objective decision-making and match recruiting practices with company objectives, analytical tools are increasingly being used to drive modern recruitment procedures. Digital technologies enable consistent and systematic candidate evaluation, which contributes to more informed and equitable decision-making. Technology also facilitates communication, resource coordination, and candidate engagement during the hiring process. As companies compete to attract and retain personnel in a competitive market, reliance on data-driven tactics and digital platforms becomes not just advantageous but also essential. This shift is indicative of a broader trend in HRM, where intuition is being supported more and more by quantitative insights and systematic processes.

With a focus on the attraction, screening, and selection stages, this article discusses the latest technological advancements in employee recruitment and selection. It highlights the importance of applicant tracking systems, internet recruiting, cyber vetting, and cutting-edge selection techniques including asynchronous interviews and gamification assessments in the hiring process. It also examines the significance of applicant replies in the selection process.[3] (Nikolaou, I., 2021).

A study investigated the effect of digital technologies in improving the recruiting process, with a particular focus on e-recruitment. This developing phenomena includes social network applicant identification, chatbot-based recruiting and job interviews, and artificial intelligence matching. It is especially valuable for ethical enterprises looking for talented personnel with shared ideals. The study uses grounded theory, participant observation, and qualitative data collection to analyze and assess a number of recruitment technologies, including Randstad.tech's big data analytic matching engine, LinkedIn, Udacity, Reveal, and TextRecruit's Ari. The report finishes with managerial tips to help recruiters implement e-recruitment.[4] (Allal-Chérif, et.al 2021)

At the moment, machine learning is one of the most promising technologies. Machine

learning has become more and more popular, particularly in the past ten years. However, many people are unaware that machine learning was developed as a result of the first attempts to teach computers to recall their own activities in 1952.

Without explicit programming, machine learning (ML) is an artificial intelligence (AI) application that can learn from mistakes and get better over time. It does this by using data—such as text, statistics, or images-to learn.

There are several use cases in recruiting where Machine Learning may be applied. The recruiting process is heavily reliant on data collection: discovering profiles, screening and rating them, analyzing prospects, and so on.

Every stage in the recruiting cycle has the potential to be heavily data-driven, which is why use cases may be found throughout.

Machine Learning is an important tool in recruiting, especially during the sourcing and screening stages. It examines applicant profiles and social media activity to learn about their history, abilities, and interests. This information is then compared to the hiring company's employment criteria. Machine Learning has an edge in terms of speed, volume, and statistical correctness when matching profile and activity information to job and corporate information. Standardizing talent data is difficult owing to differences in industry hierarchies and job definitions.

1.7.1 Assessing candidates

Candidate assessments are used to evaluate a candidate's ability. Most assessments focus on skills, but there are other methods that measure personality and intellect. In certain circumstances, using Machine Learning to help candidate evaluations makes sense since measuring talent, personality, or IQ can be far more difficult than a simple classical or fuzzy logic method.

Most Machine Learning candidate evaluation systems attempt to contextualise talents and personality attributes to the work that must be done. That's a difficult assignment because every profession has its own set of criteria, and each organisation has its own culture and requires for certain expertise

1.7.2 Interview automation

It is becoming more common, with candidates speaking to robots throughout the first interview process. While these bots aren't flawless, their efficiency and structure make them popular among organisations like HireVue and myInterview. According to MIT Technology Review, some bots graded English ability in German as a 6 out of 9. Despite their limitations, the efficiency and structure of these interviewing techniques make them useful for businesses.

Companies are using chatbots to engage applicants and pique their interest in their organisation and position. These technologies replicate real-life conversations by understanding and responding to cues and enquiries using Natural Language Processing (NLP). Mya, Olivia, and Jobpal are examples of candidate chatbots that learn by interacting with humans.

1.7.3 Using programmatic job advertising

The distribution of job advertisements over many platforms, including job boards and social media, with the goal of reaching the most suitable candidates is known as programmatic job advertising. Utilizing analytics, learning, and data, the goal is to streamline, automate, and maximize the distribution of job postings. Some programmatic job advertisement systems use classical reasoning, while others use Machine Learning processes to analyze (and change) the desirability of candidates for certain job opportunities. PandoIQ and Smart Dreamers are two examples of firms in this field.[4]

1.7 Problem Statement

Many educational institutions' current campus recruiting processes place a strong emphasis on academic performance and standardized testing. While these indicators might provide a elementary understanding of a candidate's talents, they sometimes overlook the broader spectrum of capabilities that today's sectors require—such as soft skills, cultural fit, flexibility, and the ability to solve issues on the fly. As a result, many exceptional students who excel in extracurricular activities, hands-on experiences, and specialized talents may be missed.

This restrictive method of evaluating applicants not only lowers opportunities for qualified individuals, but also creates a gap between what industries demand and what graduates can

really give. It is evident that we want a more inclusive, data-driven, and adaptable recruitment method that considers multiple aspects of a candidate's profile, fostering fairness, transparency, and a better fit with the changing demands of the workforce.

1.8 Purposes of the Research and Scope

The purpose of this analysis is to upgrade the company hiring process by putting in place a systematic, data-driven methodology that guarantees recruiters to choose applicants fairly and effectively as per their requirement. This is accomplished by luring in high-potential applicants by taking into account evaluation standards that go beyond academic performance.

The research aims to synchronize recruitment strategies with industry demands, making suggestions based on candidates' competences and skills. Besides this, it focuses on building a holistic recruitment architecture that integrates co-curricular efforts and in-app assessments. Transparency in the process of selection remains a top focus area with a goal of developing improved talent detection and increased results in hiring.

This research aims to revolutionize the traditional campus recruitment process by adopting a data driven and comprehensive approach. The goal is to close the gap between academic performance and industry expectations by using various evaluation criteria, such as co-curricular achievements, real-time assessments from recruiters, and skill-based recommendations.

Scope

- **Candidate Sourcing:** Finding high-quality candidates through a systematic selection process.
- **Industry Alignment:** Making sure recruitment strategies keep pace with changing industry needs.
- **Skill-Based Recommendations:** Utilizing data analytics to connect candidates with appropriate job roles.
- **Holistic Recruitment:** Considering both academic and non-academic factors for a

thorough evaluation.

- **Transparency and Fairness:** Creating a standardized and impartial selection framework to enhance talent recognition.

1.9 Significance of the Study

In order to better predict and improve the selection of applicants for campus placements, how might recruiters utilize machine learning algorithms to examine students' educational, technical, and personal characteristics? This research presents a recruitment framework that imprints data and AI to restructure the conventional campus placement process. The key contributions of the study are detailed as follows:

- **Ethical Hiring & Bias Reduction System :**

Encourages a diverse and merit-based hiring process, ensuring equal opportunities for all students. Minimizes recruitment bias by applying an objective, data-driven selection framework.

- **Smart Internship-to-Job Pipeline :**

Allows companies to recruit proven talent by assessing candidates through real-world projects instead of relying solely on standardized tests.

- **AI-Powered Adaptive Assessment Engine :**

Continuously updates assessment criteria to match changing industry needs, ensuring students are evaluated on both current and future skill requirements.

- **Role-Specific Dynamic Scoring Model :**

Guarantees that companies receive candidates with the appropriate mix of skills tailored to specific job roles, rather than just generic test scores.

- **Better Alignment between Education and Industry :**

Students receive training in skills that are directly relevant to what industries need. Colleges have the ability to update their curricula based on immediate feedback from the industry.

- **Promotes Lifelong Learning :**

Students are encouraged to continuously enhance their skills in response to changing industry demands.

- **Encourages Holistic Development :**

There is a strong emphasis on soft skills, cultural fit, and problem-solving, resulting in well-rounded graduates.

- **Long-Term Workforce Stability :**

By ensuring a better match between candidates and job roles, it decreases turnover and contributes to a more stable and motivated workforce.

1.10 Summary

Technology also facilitates communication, resource coordination, and candidate engagement during the hiring process. As companies compete to attract and retain personnel in a competitive market, reliance on data-driven tactics and digital platforms becomes not just advantageous but also essential. This shift is indicative of a broader trend in HRM, where intuition is being supported more and more by quantitative insights and systematic processes.

In contemporary hiring, the combination of data and technology has produced previously unheard-of levels of precision, effectiveness, and strategic insight. Throughout the hiring process, digital technology facilitates communication, resource coordination, and candidate involvement by enabling consistent and methodical candidate evaluation. In order to forecast hiring, machine learning (ML) is essential, especially in the sourcing and screening phases. By putting in place a data-based system that guarantees fair and effective candidate selection, this study seeks to improve the campus hiring procedure. Nevertheless, there are drawbacks, such as the requirement for more precise and trustworthy data, strong algorithms and easily accessible machine learning tools or methods.

CHAPTER – 2

MACHINE LEARNING TECHNIQUES

2.1 Brief on Machine Learning Algorithms

Machine learning, a dynamic branch of artificial intelligence, focuses on building systems capable of learning from data without being explicitly programmed. These systems can identify patterns, process large datasets, make informed decisions, and address complex problems. With the growing scale and intricacy of data, machine learning is becoming increasingly vital in numerous practical applications.

Broadly, machine learning methods can be categorized into different types depending on how the learning process is structured or evolves. Depending on their learning potential, ML techniques may be divided into a variety of groupings. The input-output pairs of the information that guide the learning process are used by supervised learning models to work on known data. When fitting models for regression and classification, this technique is commonly used. In unsupervised learning, unlabeled data sets are employed. These models, which are accessed through clustering analysis, seek to identify underlying structures or natural groupings within the data.

2.1.1 Logistic Regression

When the response variable follows a two-level classification (binary), logistic regression is the most effective regression method to use. Like other regression analyses, it is a predictive analysis. It is employed to describe data and the relationship between one or more dependent binary variables and one or more independent variables that are nominal, ordinal, interval, or ratio-level. This type of statistical model, often known as a logit model, is commonly used in categorization and predictive analytics. Since the result is a probability, the dependent variable has a range of 0 to 1. The chances in logistic regression undergo a logit transformation, which is calculated by dividing the likelihood of success by the probability of failure.

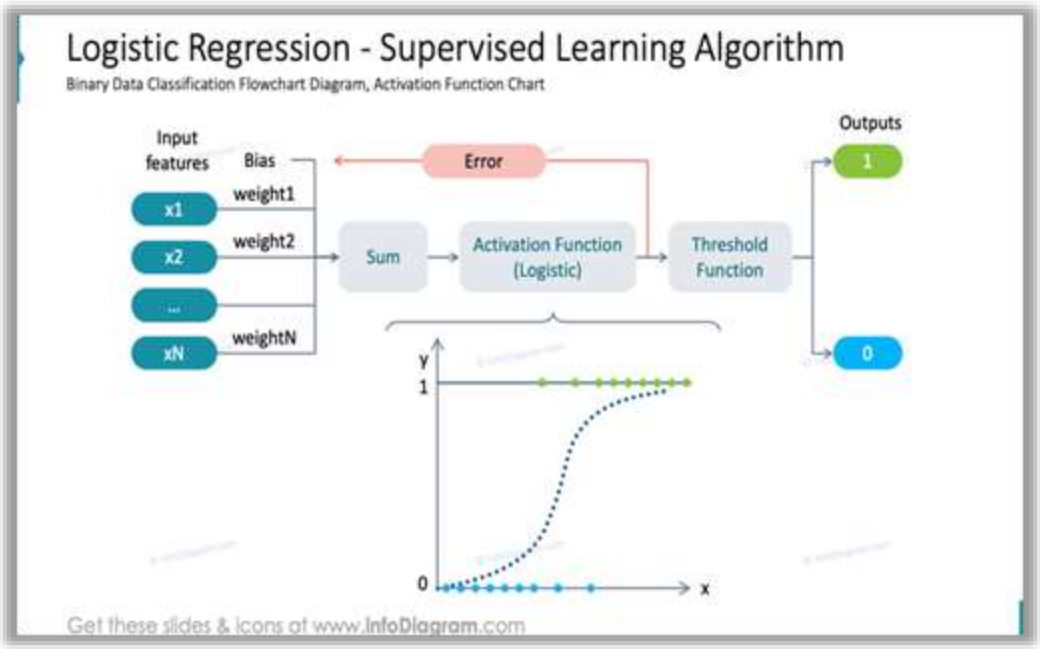

Figure 2.1 Logistic algorithm

Benefits of the Logistic regression

It is simple to implement: Training and testing are two efficient ways to implement a machine learning model. Patterns in the input data (picture) are found and linked to an output (label) during the training phase. Using a regression strategy to train a logistic model doesn't require more processing power. Thus, compared to other machine learning techniques, logistic regression is simpler to use, understand, and train.

Perfect for linearly separable datasets: A linearly separable dataset is a graph where the two data classes are separated by a straight line. In logistic regression, there are only two possible values for the y variable. Thus, it is possible to efficiently classify data into two different groups by employing linearly separable data**.**

Odds, log odds and odds ratio

The logit function, which is the basis of logistic regression, is the log of the ratio of the probabilities. We use odds instead of a linear function to describe probabilities since they can only be described by a range of 0 and 1. Despite the fact that odds and probability both indicate the likelihood of an event, they have different meanings: The possibility that an event will occur out of all possible possibilities is measured by probability.

Let p(x) represent the probability of a particular outcome. The corresponding odds ratio can then be expressed as:

$$odds(x) = \frac{p(x)}{1 - p(x)}$$

Log odds

However, the odds are not perfectly balanced, yet approximately equal to 1. For example, odds of 0.5 and 2 indicates "twice as likely" and "half as likely," respectively, on a whole diverse scale of values . This mismatch is corrected by calculating the logarithm of the odds, which converts unbounded odds scale from 0 to infinity into a continuous real-valued range spanning from negative to positive infinity. The log-odds, also known as the logit, form the basis of the logistic regression model.

It is defined mathematically as: $log((p(x))/(1 - p(x)))$

This transformation allows the log-odds to be expressed as a linear function of the input features, enabling logistic regression to model the relationship between predictors and the probability of an outcome.

$$\log((p(x))/(1-p(x))) = \alpha_0 + \alpha_1 x$$

Then, to return to odds, we can exponentiate both sides:

$$((p(x))/(1-p(x))) = \frac{e^{\alpha_0+\alpha_1 x}}{1+e^{\alpha_0+\alpha_1 x}}$$

Therefore,

$$P(x) = \frac{e^{\alpha_0+\alpha_1 x}}{1+ e^{\alpha_0+\alpha_1 x}}$$

Despite employing a linear function beneath to characterize the probabilities, this transformation enables logistic regression to get valid results.

Odds ratio

In conclusion, let us present the odds ratio, a notion that aids in the interpretation of the impact of model coefficients. If the attribute value $\alpha_1$grows by single unit, the odds ratio indicates how the odds change.

$$\text{Odds}(\alpha) = e^{\alpha_0+\alpha_1 x}$$

$$\text{Odds}(\alpha + 1) = e^{\alpha_0+\alpha_1 (x+1)}$$

This indicates that the probabilities are multiplied by α_1 for each unit increase in x. The odds ratio is this multiplier.

If $\alpha_1 > 1$ then ratio of the odds increases

If $\alpha_1 < 1$ then ratio of the odds decreases

If $\alpha_1 = 1$ then ratio of the odds is zero

The interpretability of logistic regression is attributed to the odds ratio, which provides information on how the odds of an occurrence vary with inputs.

2.1.2 Decision Tree

Decision trees are a non-parametric supervised learning method that may be applied to both regression and classification issues. It is a hierarchical tree with a root node, branches, internal nodes, and leaf nodes.

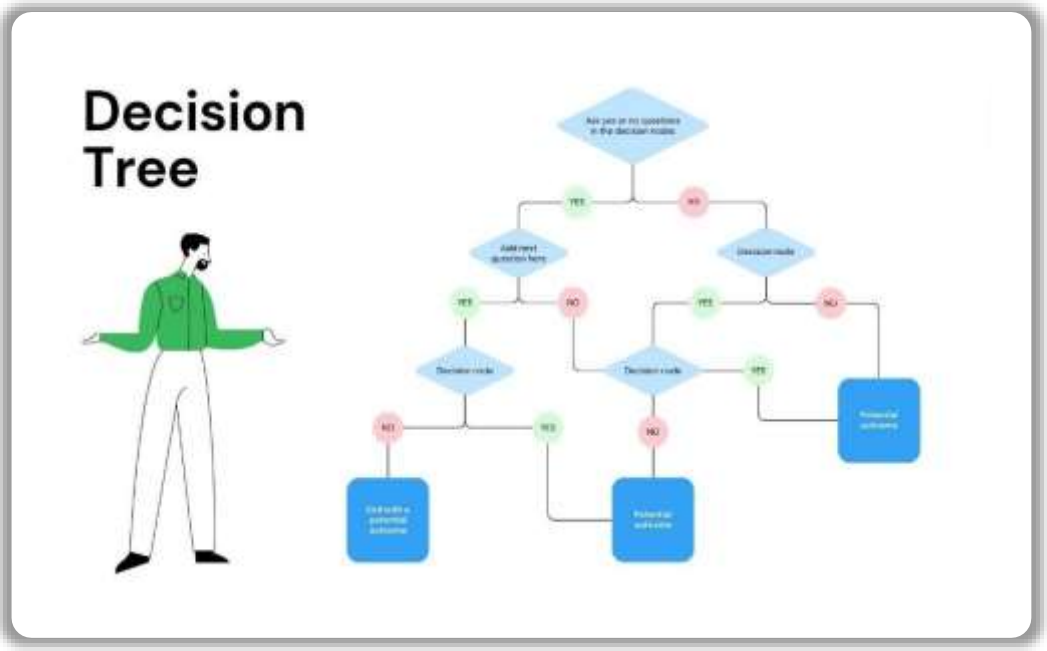

Figure 2.2 Decision Tree algorithm

Tree models use discrete target variables to classify or categorize items in classification problems. Tree models, on the other hand, are used to predict results for unseen data in regression situations where the target variable assumes continuous values (real numbers).
In machine learning and data mining, decision trees and predictive modeling approaches are commonly employed. The model makes appropriate judgments about the sample's goal value (represented by leaves) based on observations of the sample population (represented by branches).

Splitting characteristics.

In a top-down greedy approach, decision trees consider numerous parameters to determine the appropriate feature split. Greedy approaches separate all points in the same decision area, and subsequent splits are done methodically. The subsequent branch (sub-tree) has a higher metric value than the preceding tree.

Regression problem

For every data point in a decision zone, the metric is the sum of the squared variances between the mean response and the observation (target class). In order to lower the residual sum of squares, the feature splits are chosen. Therefore, feature splits are performed so that the feature value either reduces the residual sum of squares or increases information gain. For every subsequent best split, the procedure is repeated.

2.1.3 Random Forest

An algorithm that is adaptable and simple to operate, typically delivers strong results, even without extensive hyperparameter tuning. Its ability to handle both classification and regression tasks contributes to its widespread popularity.

As a supervised learning technique, Random Forest builds an ensemble of decision trees, commonly trained using the bagging approach. This technique relies on the thought that combining multiple models can lead to improved overall performance.

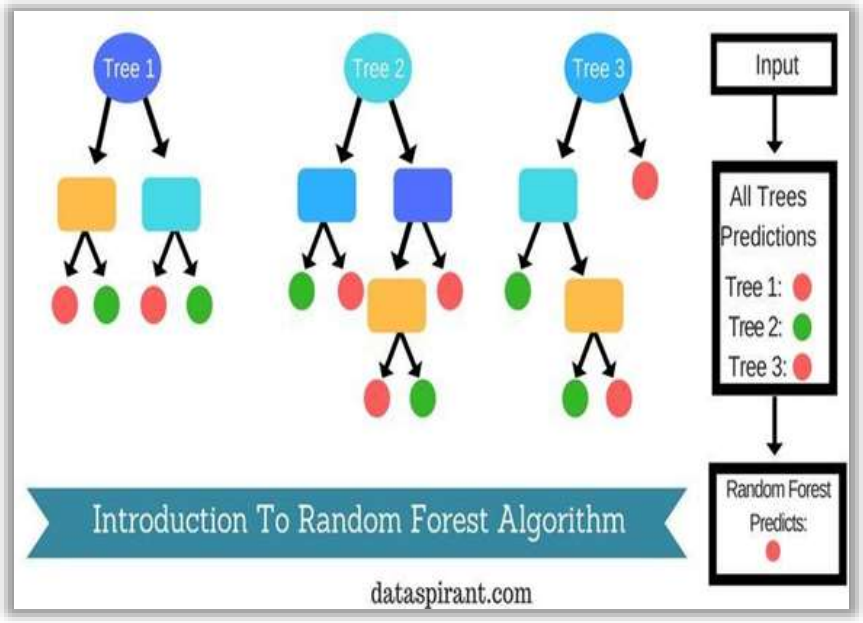

Figure 2.3 Random Forest Algorithm

Among random forests' primary characteristics are the following:

Random forests are utilized in supervised machine learning when a tagged objective variable is present.

Random forests are useful for issues that involve regression (numerical goal variable) and classification (categorical target variable).

Random forests are an ensemble approach that use predictions from several models.

The smaller models of the overall random forest ensemble are decision trees. Ensemble learning methods like bagging and boosting combine predictions from classifiers like decision trees to identify common outcomes. In 1996, Leo Breiman invented bagging. It allows for the selection of several data points by choosing a random sample of data with replacement. These models are trained independently, and the average or majority of predictions produce a more accurate estimate, depending on whether the objective (e.g., regression or classification) is to reduce variance in noisy datasets.

The random forest methodology is an extension of the bagging approach that creates an uncorrelated forest of decision trees by combining feature randomization and bagging. Often referred to as "the random subspace method" or feature bagging, feature randomization reduces correlation between decision trees by randomly sampling features. This is the main way that decision trees and random forests differ from one another. While

decision trees consider every possible feature split, random forests only select a subset of the potential feature splits.

2.1.4 Boosting Algorithms (e.g., AdaBoost, XGBoost)

Three main boosting algorithms are Gradient boost, Adaboost, and XGBoost.

Adaboost, the first adaptive boosting method, employs decision stumps as weak learners. It penalizes wrongly categorized points by increasing their weights after each prediction round, learning from earlier mistakes. The final forecast is a weighted majority vote. AdaBoost uses a large number of "weak learners" to create classifications. The weakest learners are virtually always stumps. Some stumps have more say in categorization than others. Each stump is created by taking the preceding stump's faults into consideration.

Gradient Boosting builds models sequentially, with each weak learner minimizing the residual error of the preceding one by gradient descent. Unlike AdaBoost, which adjusts sample weights, Gradient Boosting decreases error directly by optimizing a loss function.

Tianqi Chen of the University of Washington created XGBoost, an enhanced gradient boosting technique that turns weak learner trees into strong learners by adding residuals, using the same overall architecture. The library supports C++, Python, R, Java, Scala, and Julia. XGBoost's cache-aware prefetching technique reduces runtime for big datasets. On a single computer, the library may outperform rival frameworks by more than ten times. XGBoost is a scalable tree boosting system that can process billions of samples with less resource consumption due to its remarkable speed.

XGBoost integrates regularization within the learning objective, in contrast to standard gradient boosting. Data can also be regularized by hyperparameter modification. The package uses the built-in regularization of XGBoost to outperform the standard scikit-learn gradient boosting algorithm.

Hyper-parameter tweaking

To maximize the model's performance, we can experiment with various hyperparameters. The process of optimizing a machine learning algorithm's hyperparameters is known as hyperparameter tuning. The best hyperparameters may be found by iterating over a

dictionary of possible hyperparameter combinations using grid search and cross-validation techniques.

Selected hyperparameters for gradient-boosted trees in XGBoost

In gradient boosting, learning rate—also called "step size" or "shrinkage"—is the most crucial hyper-parameter. In the XG-Boost library, it is called "eta" and should have a value between 0 and 1, with 0.36 acting as the default. The learning rate determines how quickly the boosting algorithm learns new information with each iteration. A smaller eta number indicates slower learning since it reduces the contribution of each tree in the ensemble, preventing overfitting. Higher eta numbers accelerate learning, but if they are not properly controlled, they can also lead to overfitting.

The number of trees that should be produced in the ensemble depends on the hyperparameter of the estimator. The model progressively learns to fix the mistakes of the trees that came before it, adding a new tree to the ensemble with each boosting round. The complexity of the model is controlled by N estimators, which also affects how long training takes and how better the model generalizes to new data. Increasing the number of N estimators in a model frequently improves its complexity and enables it to detect more nuanced patterns in the data. However, too many trees could lead to overfitting. The learning rate often falls as the number of estimators rises.

Gamma, often referred to as the least loss reduction parameter or the Lagrange multiplier, establishes the minimum loss reduction required to divide a tree's leaf node again. A smaller value means that XGBoost completes training more quickly but might not find the best answer; a greater value means that XGBoost keeps training longer and might find better answers but is more likely to overfit. For gamma, there is no upper limit. In XGBoost, 0 is the default value, and any amount above 10 is regarded as excessive.

Max-depth specifies the maximum depth that each tree in the boosting process can reach during training. The depth of a tree is the number of layers or splits that divide its root node from its leaf nodes. Increasing this value will make the model more intricate and more likely to overfit. The default max-depth of 6 for XGBoost indicates that each tree in the model can grow to a maximum depth of six levels.

2.1.5 Support Vector Machine (SVM)

To find the optimal hyperplane that maximizes the margin between data points from opposing classes, we apply support vector machines (SVMs) in classification problems. If the hyperplane is a line in 2D or a plane in n dimensions, it relies on how many attributes are present in the input data. By widening the distance from one point to another, the algorithm can establish the most effective classification boundary and generate precise classification predictions. The ideal hyperplane is crossed by support vectors. In machine learning, SVMs are frequently used to carry out both linear and nonlinear classification tasks. In order to transform data into higher-dimensional space for linear separation, the "kernel trick" involves using kernel functions.

Types of SVM Classifiers

Linear SVMs

When working with data that is linearly separable—that is, when dividing the data into distinct groups—linear SVMs don't need any adjustments. The quadratic optimization issue contains decision boundaries and support vectors that can be visualized like a street, as described by Professor Patrick Winston from MIT, who likens it to "finding the widest street that fits between the classes." Mathematically, this separating hyperplane is represented as:

$$Wx+b = 0$$

Where b is the bias term, w is the weight vector, and x is the input vector.

The maximum distance between classes, or the margin, can be calculated using either the soft-margin classification technique or the hard-margin classification method. Professor Hinton said that the data points would be properly isolated outside of the support vectors, or "off the street." , if we use hard-margin SVMs. This is stated using the following formula:

$$(wx_i + b)y_i \geq a$$

The formula max ɣ= a / ||w|| is used to maximize the margin, where an is the margin projected onto w.

Slack variables ('ξ') are used in soft-margin classification to accommodate misclassification. The margin is controlled by the hyperparameter C; if value of C is

greater than it reduces the margin for less classification error, while having a C value smaller, it increases the margin for greater misclassified count.

Nonlinear SVM

Since many real-world data points cannot be linearly separated, nonlinear SVMs are helpful. To enable the training data that can be separated by a linear boundary, pre-processing techniques used to convert it into a higher-dimensional feature space. However, the additional computational cost and risk of overfitting might make more dimensional spaces more challenging. Using the "kernel trick" reduces some of the complexity and boosts computing performance by substituting a comparable kernel function for dot product calculations. A variety of kernel types are available for use in data classification.

Data Preparation and Model Optimization

Splitting data

The dataset is divided into training and testing sets for the Support Vector Machine (SVM) classifier, just as other machine learning methods. This reduces the danger of overfitting by ensuring that the model is trained on one set of data and assessed on another, while also promoting more generalization. Notably, this stage usually happens after considerable exploratory data analysis (EDA), which is highly recommended for model creation but not necessary. Model performance and reliability are improved by EDA, which makes it simpler to identify missing values, outliers, and a general grasp of the dataset's structure.

Hyperparameter Tuning

After the data has been appropriately separated, hyperparameter tweaking is an essential next step. Optimizing parameters such as the kernel type (e.g., linear, polynomial, radial basis function), the regularization parameter (C), and the kernel coefficient (gamma) is necessary to increase the performance of the classification model. When the SVM model is properly tuned, it finds the ideal balance between bias and variance, producing predictions that are more reliable and accurate. For this, methods like grid search and randomized search are frequently used, frequently in conjunction with cross-validation.

2.1.6 KNN (K- Nearest Neighbor) Algorithm

A ML method which determines how close two data points are to one another is k-nearest neighbors (KNN) algorithm. For both classification and regression issues, it is a popular option because of its accessibility and simplicity of usage.

The KNN algorithm is a supervised learning method, which means it uses previously learned training datasets. This labeled input data is used to train a function that produces an appropriate output when supplied with more unlabeled data.

The program then uses the distances between the input data point and its K nearest neighbors. The method chooses the most common class label among the K neighbors in the classification instance to determine the predicted label for the input data point. Regression predicts the value for the input data point by utilizing the weighted average or average of the target values of the K neighbors.

Regression and classification issues may be fixed with the KNN approach. However, it is most frequently used for industrial classification issues. When assessing any approach, we usually consider three crucial factors:

1. Simplicity of output interpretation.

2. Time spent on calculations

3. The Ability to Predict

Working of KNN

The most popular distance in regression problems, which average the k closest neighbors to predict a classification, is the Euclidean distance. The KNN algorithm, which belongs to the family of lazy learning algorithms, saves a training dataset and does calculations while predicting or classifying data. Data scientists continue to use it for financial market forecasting, pattern recognition, intrusion detection, data mining, and simple recommendation systems.

To identify which data points are in close proximity to a specific query point, calculating the proximity between the query point and other data instances. These distance metrics help to create decision boundaries by segmenting query points into discrete zones.

Decision boundaries are commonly represented using voronoi diagrams.

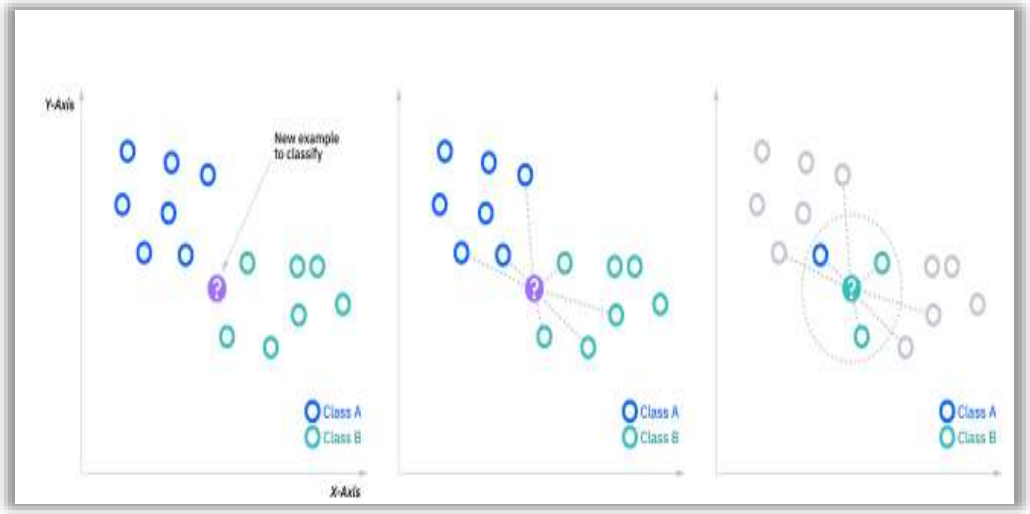

Figure 2.4 KNN Process (Source: ibm.com)

Only real-valued vectors may be measured using the most used distance metric, the Euclidean distance (p=2). To compute the straight path between the query and another position, the following formula is applied.

$$d(\alpha, \beta) = \sqrt{\sum (\beta_i - \alpha_i)^2}$$

Another well-liked distance metric that calculates the absolute value between two places is the Manhattan distance (p=1). Since it is frequently shown with a grid that shows how the approach one takes to travel from one place to different place via city streets, it is sometimes called as city block distance or taxicab distance.

$$d(\alpha, \beta) = (\sum |\alpha_i - \beta_i|$$

The k parameter of the k-NN algorithm determines how many neighbors are looked at in order to categorize a query location. A compromise between overfitting and underfitting may be achieved by determining k; larger values lead to greater bias and less variance, while smaller values result in high variance and less bias. The value of k will be determined by the input data; an odd number is recommended to avoid classification ties. Cross-validation techniques may help determine the optimal k for the dataset.

Various computing KNN distance measurements

For the kNN method to function, it is essential to determine the distance between the query point and the other data points. Determining distance measures enables the establishment of decision boundaries. These boundaries establish different zones of data points.

Many techniques are used to compute distance:

Euclidean distance, the most popular distance metric, determines how long a straight line would be from the query location to the other place.

Another popular distance metric is the Manhattan distance, which determines the absolute value between two locations. From point A (your inquiry point) to point B (the point being measured), how do we get there? This is shown on a grid and is referred to as taxicab geometry.

A technique for figuring out whether Boolean or string vectors do not match is the overlap metric, sometimes referred to as the Hamming distance. In other words, it determines how far apart two strings of the same length are. It is very useful for error correction and code identification.

2.2 Summary

Machine learning is a subfield of artificial intelligence that focuses on developing systems that can learn from data and get better at what they do over time. When dealing with dependent variables that are dichotomous, logistic regression provides useful information. Random forests and decision trees are examples of non-parametric supervised learning methods used for classification and regression. For regression and classification, Support Vector Machines (SVMs) are applied, even though the K-nearest Neighbor (KNN) approach chooses the most common class label among K neighbors to forecast labels.

CHAPTER – 3

MOTIVATION

3.1 Motivation

The study aims to increase the equity and effectiveness of campus placement systems. Using the following themes, the rationale is examined:

a. Importance of the Campus placement

A key factor in closing the gap between academia and business is campus placement. It helps businesses hire qualified recent graduates and gives students early career opportunities.

b. Challenges in the traditional Placement Processes

Standardized skill tests and academic performance are frequently major factors in traditional placement procedures. Due to these approaches' limited capacity to evaluate practical skills or real-world problem-solving abilities, student readiness and industry expectations frequently diverge.

c. Machine Learning's Function in Placement Prediction

For traditional methods, machine learning (ML) models offer data-driven alternatives. Both supervised learning algorithms, such as KNN, Decision Trees, Random Forests, and Logistic Regression, and deep learning models, such as Artificial Neural Networks (ANNs), can be used. It is possible to predict eligibility using other probabilistic methods, including Bayesian classifiers.

d. Key Parameters for the company’s placement prediction

CGPA, subject-specific academic performance, technical and programming skills, internships, project experience, aptitude scores, problem-solving abilities, communication and teamwork skills, industry-relevant certifications, and involvement in extracurricular activities are all significant factors that affect placement prediction.

e. Benefits for Institutions and Companies

Academic institutions can improve student employability by customising training programs with the aid of this predictive approach. Additionally, it helps recruiters make objective, well-informed decisions, which results in hiring procedures that are more fair and effective.

f. Enhancement of the placement Strategies

The model can assist in finding hitherto unnoticed predictors that employers value, make it easier to match academic programs with industry demands, and provide students with individualized career counselling and skill-development techniques.

3.2 Challenges

a. Data Quality and Availability [5] [6]

- Incomplete or inconsistent student records, including academic performance, afterschool activities, and work experience, have an impact on how predictions can be made.
- Colleges using different ways to gather information make it hard to create a standard approach.
- What's more, worries about keeping personal details private make it tough to get full student profiles, which makes it harder to train models well.

b. Bias in Machine Learning Models [5]

- Old placement data might favour students from certain backgrounds, schools, or areas. When AI systems learn from biased info, they can widen existing gaps putting some students at a disadvantage.
- To ensure fair predictions, you need to use ways to cut down bias. However, putting these ways into action isn't a walk in the park.

c. Dynamic Industry Requirements [7]

- The job market trends are constantly shifting, which means that models need to be updated regularly.
- Skills that are in demand, such as AI and cloud computing, change over time, rendering older training data less useful.
- Moreover, various industries and companies have their own specific hiring criteria, making it difficult to create a one-size-fits-all prediction model.

d. Algorithm Selection and Performance [4] [7]

- Not every machine learning algorithm is the best fit for every dataset; the choice depends on the specific traits of the data.
- Certain models, like deep learning, demand significant computing resources, which can lead to high implementation costs.
- It can be difficult to strike a balance between computing efficiency, interpretability, and forecast accuracy.

e. Interpretability and Trust in Predictions [7] [5]

- Neural networks and ensemble techniques are examples of complex machine learning models that frequently operates like opaque systems, making it challenging to understand the reasoning behind their decisions.
- This lack of clarity can lead to scepticism among recruiters and students regarding placement predictions.
- While creating explainable solutions can introduce additional complexity, it is essential for gaining acceptance and trust in these technologies.

f. Scalability and Customization [7] [10]

Placement criteria differ among universities, which complicates the application of a uniform model across institutions. Adapting to various student profiles and job markets necessitates flexible machine learning frameworks. Additionally, scaling models to handle large datasets raises demands for computational power and resources.

g. Real-World Implementation Challenges [11] [5]

Many institutions want to use AI techniques for recruiting the candidates rather than using the conventional approach. Adoption of these advancements may be more challenging. It is possible that such companies may not be able to get infrastructure facilities such as computers with high computational power and the costs associated with their maintenance. The maintenance workload is also increased because models must be regularly retrained to stay up with shifting employment trends.

h. Mismatch between Predictions and Actual Hiring Decisions [9] [7] [8]

- The AI powered placement strategy is hesitantly adopted by many universities and the placement agencies as there may be a chance of the mismatch of the candidates and they opt traditional techniques over the AI powered.
- The universities or the recruiting agencies may face the problem of the Lack of processing capacity and machine learning expertise. These add limitations to their requirements.
- In order to keep up with changing job trends, models need to be retrained on a regular basis, which increases the maintenance workload.

CHAPTER – 4

LITERATURE REVIEW

4.1 Introduction

The study's objective is to review and assess recent research on campus placement prediction and the application of machine learning (ML) methods in educational data mining. In Indian institutions of higher learning, campus placements contribute a significant part in identifying students' career possibilities. As the labour market gets more competitive, researchers and educational institutions are examining the many factors that affect placement success. Simultaneously, new opportunities for analyzing vast amounts of student data to predict placement success have been made possible by the advancements in machine learning and predictive analytics.

4.2 Machine Learning's Function in Campus Placement Predictive Analytics

Byagar et al. [9] looked at a variety of techniques, factors that affect placement, and the benefits and drawbacks of employing various algorithms. Three machine learning algorithms—Random Forest, Decision Trees, and Naïve Bayes—were used to estimate student placements. These algorithms' accuracy was evaluated.

In order to ascertain the likelihood of current students being put on campus, Shahane, Priyanka et al. [10] used four machine learning algorithms: Random Forest, K-Nearest Neighbours, Decision Tree, and Logistic Regression. They did this by analyzing last year's student placement data. Help in this area may have a major influence on an institute's ability to place students, according to their findings.

Priya Swaminarayan and M.R. [11] aimed to develop a prediction model to forecast a student’s chances of securing placement. They examined historical data from the previous academic year, estimated placement probabilities for current students, and supported

efforts to improve placement success rates. Four classification methods were used in the study's suggested recommendation system: Support Vector Machine (SVM), Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN). The dataset's characteristics were used to evaluate these algorithms' efficacy. This study assisted in identifying students who showed promise academically, allowing them to improve their social and technical abilities to improve their chances of getting placed.

Shubham Khandale and et al. [12] given that placement at respectable MNCs is essential to their reputation and annual admittance, academic enthusiasts' main objective is to secure employment there. An institute's training and placement office (TPO) can benefit from a method for forecasting student placements. Machine learning algorithms are able to predict future placement pupils and extract knowledge from past ones. Training data is sourced from the same organization, and suitable pre-processing techniques are applied. Applying sophisticated techniques like bagging, boosting, and voting classifiers produced 78% in XGBoost and 78% in AdaBoost Classifier.

S. Rao et.al, [13] proposed a system for using past academic achievement and extracurricular activities to forecast suitable placement categories (dream businesses, super dream firms, and mass recruiter corporations). The algorithm also detects talent for future recruitments, which aids in placement preparation. Real-time experimental findings and performance metrics are provided for model validation, with the purpose of meeting outcome-based education (OBE) milestones in educational institutions. This ensures that students will have more options for placement and professional growth.

Campus placement was crucial for both students and educational institutions, as admissions often depended on placement outcomes. P. Jain et al., [14] by anticipating assignments in advance, the study sought to lessen the burden on training and placement cells. The experiment included machine learning models, including k-NN classification, Gaussian Naïve Bayes, Decision Tree classification, SVM classification, Random Forest classification, Logistic Regression classification, and Ensemble Voting classification. There was less work for the training and placement divisions because the voting classifier was the most accurate of them.

In order to predict students' placement in the IT business based on their academic performance in a range of disciplines, Maurya, L. S. et al. [15] created supervised machine learning classifiers. Heat maps, classification reports, confusion matrices, accuracy scores,

and other indicators were utilized to evaluate the classifiers. Techniques including Gaussian Naïve Bayes, Random Forest, Decision Trees, K-Nearest Neighbor, Logistic Regression, Random Forest, Gaussian Gradient Descent, and Neural Networks were employed. Every classifier was tested on new datasets, and measures for accuracy, recall, f1-score, and support were used to assess the outcomes.

To solve unconstrained issues, a novel optimization technique called the Adaptive Fuzzy Campus Placement Optimisation Algorithm (AFCPOA) was developed. Fuzzy models were used to describe written tests and interview procedures by A. Sunil et al. [16]. Performance of the technique was evaluated on ten benchmark optimization functions and compared with alternative methods. When applied to the IEEE 33-bus radial distribution system, it performed better than earlier techniques for the optimal size and location of distributed generators.

An educational institution's reputation and yearly admissions are significantly influenced by the placement of its pupils. Institutions strive to improve their placement departments in order to solve this. An algorithm will be used in Sharma, A., & Kaur, R.'s [17] study to evaluate student data from the prior year and project the placement prospects of current students. The same organization provided the data, which was pre-processed using the proper techniques. The accuracy, precision, and recall of the proposed model were evaluated against well-known classification techniques as Random Forest and Decision Tree. The findings revealed that the suggested algorithm outperformed the other algorithms, highlighting the significance of successful placement techniques for both students and institutions

A placement predictor that calculates the likelihood that a student would be employed based on the company's needs was proposed by Jose, I. T. et al. [18]. A variety of criteria are used by the placement predictor to assess the student's competence level. At the university level, some criteria are collected, while others are obtained through assessments given on the placement management system itself.

Placement prediction is critical for institutes seeking to boost placements and admissions. Agrawal, V. S., & Kadam, S. S., [19] to determine students' eligibility for placements and their chances of passing various interview processes, such as aptitude, programming, group talks, and HR or technical interviews, the proposed technique uses a Naïve Bayes classifier. The method's use of relief feature selection techniques will improve Naïve

Bayes. Registration number, name, program, 10th and 12th grades, CGPA, standing arrears, logical thinking, numerical ability, English proficiency, and coding competence are all included in the dataset that was used.

A study by Dhingra, M.et.al [20] examined how 29 Indian higher education institutions (HEIs) students felt about campus placement programs and how important they were to employers' selection criteria. Significant differences were found among the five categories when the data was analyzed using statistical methods including average, t-test, and Garrett Ranking. Businesses believed that the most important factor affecting students' acceptance or rejection during campus placements was how well they performed during interviews. The results can help schools create a successful strategy for getting kids ready for the workforce.

Clinical rotations are critical for health students to complete work-integrated learning requirements and build professional skills in real-world situations. Smith, B. et.al [21] to save resources and increase opportunities, it is essential to analyze the elements that determine placement provision. Institutional, personal, university participation, and student competency were the four main themes identified in a scoping review of global literature published between 2000 and 2020. Placement provision depends on a number of factors, and understanding them is essential for long-term development and resource sustainability. This research lacks the perspectives and experiences of health care managers and physicians who choose not to participate in clinical placements. Future research in this field will greatly benefit from the input of these important specialists.

S. Bhoite et.al [22] developed a mechanism that will help the institute's training and placement operations by automatically forecasting student placement in the early stages of their studies. Classifying pupils and overcoming computational complexity through the deliberate construction of several models are the goals of ensemble learning. Through the use of machine learning approaches like AdaBoost classifier, K-NN, SVM, Decision Tree, Logistic Regression and Random Forest, the suggested method addresses the problem of choosing weak people. The purpose of this approach is to improve decision- process in the area of higher education.

According to Navyashree, S. L et.al [23] Technical colleges must thrive in campus placements by prioritizing and training students to perform well in interviews. A student placement prediction and recommender system is required to allow placement authorities

to forecast placements. Advances in machine learning algorithms can help institutions and students flourish. Several methods have been used to predict student placement, including decision trees, k-nearest neighbors, and support vector machines. This study examines different algorithms and offers advice on how to select an efficient approach for developing the system model based on training data. This approach can help institutions and students succeed in their academic pursuits.

According to Srinivas, K. *et.*al, [24] students have been more interested in Machine Learning (ML) algorithms for real-world prediction challenges. A number of models, including linear regression, K-neighbor regression, decision tree regression, XGBoost regression, gradient boost regression, light GBM regression, and a random tree classifier model, are used in this work to predict student placement. A basic data set is used in Phase 1 of the study, whereas an expanded data set containing additional student features is used in Phase 2. With a focus on prediction accuracy and root mean square error, a comparative performance analysis of several models is presented.

Sahu, B. L., & Tiwari, A. [25] discussed that Machine learning, a subset of data science, provides rapid and efficient solutions for big and diverse datasets. It offers fast, well-organized algorithms and data-driven models for real-time data processing. Linear Regression, a supervised machine learning method, can assist institutions in forecasting future student placement trends based on advanced placement practice test results. This allows students to identify and improve their areas of weakness, resulting in greater institutional placements. Institutions want to give golden possibilities to their students by using machine learning into their operations.

Goyal, J., & Sharma, S. [26] describes a method for developing a placement prediction model for pre-final year engineering graduate students using deep learning techniques. A placement probability predictor helps students evaluate their possibilities of getting a decent job and aids in academic preparation for the future. Data mining and deep learning were utilized to analyze previous year's student datasets, introducing several predictive models. The study describes a novel way for developing a prediction model for pre-final year engineering graduate students**.**

S. Gaurkhede et.al [27] studied that a prediction approach for campus placements employs a variety of characteristics to assess a student's skill level. This strategy improves academic performance and soft skills by using data from exams, college-level sources, and other

aspects. The purpose is to increase students' placement opportunities and reputation. Schools frequently expand their placement department to strengthen their organisation. This research looks at student placement data from the previous year to determine the chances of current students being placed on campus. Four machine learning methods, Logistic Regression, Decision Tree, Support Vector Machine, K Neighbor Classifier, and Random Forest, were evaluated for this aim.

The research by Rao, K. E. et.al [28] pursued to enhance job placement predictions for graduate students by collecting data from individuals who finished courses at various universities. The researchers gathered data on their academics, performance, families, abilities, personal information, and habits to build a synthetic data model. The model utilizes synthetic data to forecast student career outcomes, allowing universities to better prepare students for their future employment. The eXtreme Gradient Boost (XGBoost) machine learning model was tested for accuracy and precision against common classification approaches, and the findings revealed that the suggested algorithm outperformed the alternatives. This research attempts to assist students better prepare for their future occupations.

The creation to predict student placement through Machine Learning was prompted by the sharp rise in seek higher study in Turkey. Çakıt, E. et al. [29] Compared to the other approaches, the Extreme Gradient Boosting (XGBoost) algorithm delivered good result in terms of expected accuracy. A sensitivity research found that the five most productive input attributes were % of student placement at year t-1, scientific document score of university, and score of degree programme, score of faculty member / student, and percentage of student placement at year t2. University quotas, resource allocation, and the creation of new regulations can all be influenced by these projections and the results of the sensitivity analysis.

Animesh Gir and et al. [30] suggested a placement prediction system that forecasts undergraduate students' likelihood of being hired by an IT business. The author used the machine learning KNN classification method, taking into account a number of factors, including their academic achievement, that are evaluated by the business. In addition, they contrast the outcome with those of other models, such as SVM and Logistic Regression.

Marium-E-Jannat and et al. [31] suggested a method that aids the business in narrowing down the pool of applicants based on factors including CGPA, completed projects, area of

interest, research, and experience. Since finding the right candidate involves a number of steps, the creator of the Naive Bayes Classifier suggested a method that shortlists applicants and allows the parameters of the criterion to be altered as necessary. A candidate with a high CGPA may not be eligible for a position, whereas a candidate with a low CGPA may be, based on the author's findings.

Mangasuli Sheetal B and et al. [32] Suggested a system that employs fuzzy logic and the data mining techniques KNN and Fuzzy Logic to predict whether a student would be put on campus based on a variety of factors, including soft skills, extracurricular activities, and academic information. The accuracy and performance level of the output from the two algorithms are compared. According to the study of the results, KNN provides the best outcome for predicting the likelihood of placement.

Ashok MV et al. [33] suggested a model with an algorithm to forecast the percentage of students and institutions placed. The method is evaluated against neural networks, decision trees, and naive bayes and is determined to be the most effective. The data analysis, which is based on the students' factual information from the previous period, serves to raise the likelihood of placement, which in turn helps to boost admittance.

The approach that predicts the placement position for B.Tech students in five groups—Dream Company, Core Company, Mass Recruiters, Not Eligible, and Not Interested in Placements—was presented by Thangavel S. et al. [34]. To take the appropriate action, the system additionally assists the placement management staff in identifying the weaker and prospective pupils. By this system students can also put in more and extra hard work for getting selected into the companies.

Syed Ahmed and et al. [35] proposed the placement management system to predict the eligible students for the placement according to the company criteria for the placement. It helps the student from the start to prepare for the campus placement and improve their weaker area so that they can perform accordingly. The TPO will check various attributes like academic history, technical skills, aptitude, etc which are tested by the companies at the time of recruitment. Authors have used the C4.5 algorithm which is applied on company previous year data and the current criteria to initiate the result.

In addition to predicting the firm based on the historical data of the students who have previously been placed, Harinath S. et al. [36] focuses on a technique that predicts the

placement prospects of current year students in the company. Two machine learning methods—Naïve Bayes and K-nearest neighbor (KNN)—are used to make the forecast. These methods take into account a number of factors, including CGPA, technical and other abilities, and grades 10 and 12.

Manvitha P. et al. [37] proposed the system which predicted the chance of graduate students being placed by the company through the campus recruitment. Author predicts the analysis by using classification algorithms like Decision tree and Random Forest. For the placement prediction various parameters were considered like academic percentage, backlogs, credits etc. By this model authors try to improve the campus recruitment in institute as well it will help in improving student performance.

T. Aravind et al. [38] proposed a method using seven machine learning algorithms to solve the placement prediction issue and forecast the student's yearly pay using historical data. Two steps are involved in the execution of the method, which includes the following models: linear regression, K-neighbor regression, decision tree, XGBoost, gradient boosted, light GBM, and random tree classifier. The K-neighbor regression model is said to be the best when the first phase is carried out on basic data. In the second phase, additional parameters such as the programming language and the region of interest are introduced, and the best results are obtained using both the linear regression model and the K-neighbor regression model.

In their study, Kumar Satish et al. [39] used a logistic regression technique to estimate the likelihood that a student who chose the management course would be deployed on campus or not. The author selects five top institutions in Andhra Pradesh for the data set and utilizes six criteria to construct predictions: CGPA for PG and UG, specialization for PG and UG, soft skill scores, and gender. These factors are analyzed using R.

Srinivas C K et al. [40] predict the chance of being placed or not in a campus recruitment. By this model, the author proposed that it will help both student and institutional placement management for handling campus recruitment. By this they can find out the potential value and position of the student by predicting the chances of being placed which help student and institutional placement management to enhance the performance of potential students. For this they have used a Logistic regression algorithm in machine learning which includes the criteria like academic performance including technical and verbal skills and aptitude which are the part of campus selection.

G.Gautmi et al. [41] proposed a system which assists both student and placement management team to look over student potential and let them know their weak points using Linear Regression tool of machine learning and with the help of result appropriate action can be taken to improve student weaker points so more number of placements can be achieved by the institute.

The organization may save a significant amount of time and money by using the method that Reddy J. et al. [42] offered to forecast the right applicant before choosing the resume. Factors such as age, number of previous employers, experience, pay increase, and so on are taken into account when predicting. The method is carried out with the use of machine learning techniques such as Random Forest, K-Nearest Neighbor, Decision Tree, and Gaussian Naive Bayes.

Samatha, J.et.al [43] intends to use student data to forecast future student standards and placement prospects. This evaluation is based on academic talents, creativity, research, and development capabilities. A Google form survey was used to collect data from students in Chhattisgarh's computer science, engineering, and information technology departments. The research necessitates developing a dataset, which is tough owing to the difficulties in encouraging students to submit information. The dataset is then used to predict placement outcomes with a Supervised Machine Learning Algorithm (Classifie), which is one of the most efficient and straightforward algorithms for supervised labelled datasets. This project attempts to raise students' standards and develop courses and extracurricular activities accordingly.

Pune, B. [44] discussed that based on a company's requirements, a placement predictor is a tool that forecasts a student's likelihood of being hired. It evaluates a student's ability level using multiple aspects like data from universities and assessments taken through the placement management system. The predictor ensures a fair and accurate conclusion by using historical student data to accurately anticipate the student's placement.

Archana, P., et.al [45] examined the past data from students from previous year and forecasted their placement prospects to forecast placement results, the system employed two machine learning classification algorithms: K-Nearest Neighbors [KNN] and Naive Bayes Classifier. Using information from students who have already been placed, the algorithm also forecasts where current students will be put. Companies use the approach to place students at predetermined intervals so they may concentrate on honing their social

and technical skills. The dataset is used to compare various algorithms' levels of efficiency.

The use of recommender frameworks and job portal websites to recommend appropriate positions to applicants is covered by Parida, B., et.al (2022) [46]. After cleaning and filtering data using machine learning techniques like Random Forest Classifier (RFC), the framework employs optimization approaches to produce the best accurate results. The advantages of the recommender framework in career orientation are highlighted in the study, especially in geo-area-based recommendation, which assists job seekers in locating the ideal position based on their preferences. The study comes to the conclusion that, especially when it comes to career orientation, the usage of job recommender systems can greatly enhance the recommendation of appropriate employment for job searchers. All things considered, the study emphasizes how job recommender systems might improve job seekers' job search experiences.

In this paper Viram, R. et.al [47], a support vector algorithm-based machine learning-based placement prediction system is presented. Based on students' performance, communication abilities, and interest in their ideal firm, it creates a Google form. The system offers technical interest, outcomes tracking, and guidance. The algorithm creates new data to provide alternatives based on accomplishments and abilities in the event that students are unable to secure their ideal employment because of company requirements. We'll talk about the system's future significance and breadth.

In spite of their diligence and hard work, many engineering students in India are not hired, making the recruitment process extremely difficult. In order to improve student eligibility and identify critical attributes for recruiting, Mulye, V., & Newase, A. [48] proposed a prediction method utilizing data mining techniques. This approach could enhance recruitment results and assist students and institutions in improving their attributes.

In order to forecast the likelihood of student placement, this study by Vidyashreeram, N., & Muthukumaravel, A. [49] created a predictive model. It estimates placement opportunities, examines historical data from prior academic years, and suggests a recommendation mechanism. K-Nearest Neighbors, logistic regression, random forest, and support vector machine are the four machine learning classification algorithms that are employed. By identifying students with academic promise, the approach helps them develop their technical and social abilities, which increases their chances of getting placed.

The study by Pandey, A., & Maurya, L. S. [50] forecasted the post-graduation job prospects of engineering students using classification technique Criteria like the accuracy score, confusion matrix, heatmap, % accuracy score, and categorization report used to compare and evaluate the results. For the purpose of helping students which make well-informed decisions based on their academic achievement and skill set, the research attempts to discover elements that influence their decision-making process while selecting the appropriate career route using machine learning approaches.

This study by Kumar, R. S. et.al [51] introduced a prediction algorithm that uses academic and aptitude data to forecast student placements. The model employs data pre-processing techniques and evaluates data gathered from institutions. It makes independent predictions using a single supervised machine learning technique, the support vector machine algorithm. In order to increase the placement rate of schools, the model assists placement cells in concentrating on prospective students and enhancing their technical and other skills. The dataset is employed to measure the model's performance, guaranteeing precise predictions and improving the placement procedure.

While online learning systems such as MOOCs and VLEs give students freedom, they also have drawbacks, such as poor self-regulation and low engagement. This study by Adnan, M., et al. [52] proposed a predictive system to evaluate the problems experienced by kids who are in danger and provide timely support. The model uses machine learning and deep learning techniques to explain learning behavior. The strategy helps teachers to identify academically vulnerable students early in order to prevent dropouts. Time-dependent variables, degree of involvement, and assessment outcomes are all crucial components of online learning. The best results are obtained by models trained with Random Forest (RF), which have average accuracy, precision, recall, and support.

Table 4.1 Comparative Summary of Literature on Machine Learning in Predictive Analytics for Student Placement

Ref	Focus Area	Methods/ Techniques Used	Key Findings / Contributions
[9]	Factors influencing campus placement outcomes	Naïve Bayes, Random Forest, Decision Trees	Comparative accuracy analysis of classifiers; identified strong-performing models for placement prediction
[10]	Likelihood of	Logistic Regression,	Demonstrated that predictive

	current students' placement based on prior year data	Decision Tree, K-NN, Random Forest	insights can significantly improve institutional placement success
[11]	Student placement probability forecasting	K-NN, Logistic Regression, Random Forest, SVM	Developed a recommendation system to guide skill development for better placement prospects
[12]	Predicting reputable MNC placements	Bagging/Boosting/XG Boost, AdaBoost	XGBoost and AdaBoost each achieved ~78% accuracy, showing ensemble methods' benefits
[13]	Placement-category prediction for OBE framework	(Likely ANN/SVM/KNN—algorithms not explicitly listed)	Proposed a methodology for categorizing placements; supported OBE goals with metrics and real-time validation
[14]	Early-stage placement forecasting	Logistic Regression, SVM, Decision Tree, Gaussian NB, Random Forest, K-NN, Voting Ensemble	Voting classifier produced the highest accuracy, easing load on placement teams
[15]	Engineering student placement prediction	SVM, Gaussian NB, K-NN, Random Forest, Decision Tree, SGD, Logistic Regression	Compared learner performance using aptitude and coding skills; detailed model comparison provided
[16]	Optimizing placement procedures using fuzzy logic	Adaptive Fuzzy Campus Placement Optimisation Algorithm	Introduced a novel fuzzy-logic algorithm; validated using benchmark functions and IEEE 33-bus model
[17]	Institutional placement strategy evaluation	Decision Tree, Random Forest, custom algorithm	Custom method outperformed standard classifiers based on accuracy, precision, and recall

[18]	Predicting placement likelihood based on company criteria	Rule-based predictor using company-specified student features	Developed a framework to align student skills with company requirements
[19]	Placement prediction with skill + academic data	Naïve Bayes + Relief feature selection	Predicted eligibility based on CGPA, aptitude, coding; feature selection improved model
[20]	Student & employer perceptions of placement experiences	Statistical methods: average, t-tests, Garrett ranking	Employer interview performance deemed most influential; suggested enhanced preparation strategies
[21]	Determinants in health-education placement provision	Scoping literature review	Identified factors like institutional policies, student competence; emphasized resource sustainability
[22]	Early prediction for training/resource planning	Ensemble methods (LR, SVM, KNN, DT, RF, AdaBoost)	Early automated predictions assist TPOs with student tracking and planning
[23]	Student placement + recommendation systems	Decision Tree, KNN, SVM	Compared algorithm efficiency; recommended ML integration in placement processes
[24]	Two-phase predictive modeling (basic vs enhanced datasets)	Linear/ KNN/ Decision Tree regression, XGBoost, Gradient Boost, LightGBM, Random Tree	Phase-wise evaluation using RMSE and accuracy; detailed performance comparison
[25]	Placement trend prediction using	Linear Regression	Identified student weaknesses; showed potential for

	test data		improvement via test-based ML predictions
[26]	DL-based predictive model for early placement prep	Deep Learning, Data Mining	Presented novel deep-learning model aimed at early career guidance
[27]	Skill-based placement prediction	Logistic Regression, Decision Tree, SVM, KNN, Random Forest	Improved academic and soft-skill assessment to boost placement opportunities
[28]	Cross-university career outcome prediction using synthetic data	XGBoost vs other classifiers	XGBoost outperformed peers; recommended best practices for future employment prep
[29]	Student placement forecasting in Turkish HE	XGBoost + sensitivity analysis	XGBoost achieved highest accuracy; identified five key input variables
[30]	IT placement chance prediction	KNN, SVM, Logistic Regression	KNN-focused; compared against other classifiers using academic performance
[31]	Candidate shortlisting via CGPA/experience	Naïve Bayes	Dynamic thresholds enabled eligibility adjustments; CGPA wasn't always top predictor
[32]	Campus recruitment prediction using fuzzy logic	KNN, Fuzzy Logic	Combined academic and soft-skills data; KNN showed stronger results
[33]	Projecting student/institute placement rates	Decision Tree, Naïve Bayes, Neural Network	Proposed model outperformed traditional ones, positively impacting admissions
[34]	Multi-category placement forecasting	(Classification via MRC/Dream categories)	System classified students into 5 categories for targeted student support
[35]	Company-criteria	C4.5 decision tree	Early tracking of student

	based placement management		preparedness; aligned with company criteria
[36]	Predicting placement chances + company matching	Naïve Bayes, KNN	Considered academics + technical/social skills for forecasts
[37]	Graduate placement prediction	Decision Tree, Random Forest	Used multiple parameters (backlogs, credits); model aimed at institutional use
[38]	Salary prediction + placement chances	Linear/KNN/Decision Tree regression, XGBoost, Gradient Boost, LightGBM, Random Tree	Two-phase study: simple vs enriched dataset; KNN and linear regression led
[39]	Placement forecasting for management students	Logistic Regression	Analyzed PG/UG CGPA, soft skills, gender across AP colleges
[40]	University placement model for student/institution	Logistic Regression	Evaluated technical/verbal/academic skills; aided placement planning
[41]	Student-potential assessment tool	Linear Regression	Highlighted weak points; informed targeted actions to boost placements
[42]	Resume-based candidate filtering	Gaussian NB, Decision Tree, KNN, Random Forest	Focused on profile factors like experience and industry movement
[43]	Student Forecasting Model on Employability	Supervised ML (classifier unspecified)	Aimed at predicting placement; challenged by data collection; promotes course and activity adjustments
[44]	Placement	Placement predictor	Predicts student hiring

	prediction based on institutional data	using historical data	likelihood based on assessments and academic data
[45]	Placement forecasting from past student data	KNN, Naive Bayes Classifier	Compares algorithm efficiency; helps students improve social and technical skills
[46]	Job recommendation system	Random Forest, Optimization techniques	Enhances job search via geo-location based recommendations and career orientation
[47]	Placement prediction system	Support Vector Algorithm (SVM)	Assesses performance and interests; suggests alternate job options based on skills
[48]	Enhancing recruitment using student attributes	Data mining techniques	Identifies key recruitment traits; improves student eligibility and institutional outcomes
[49]	Predictive modeling of student placements	KNN, Logistic Regression, RF, SVM	Uses academic history to boost placement chances and recommend skill improvements
[50]	Post-graduation job prediction	ML algorithms with accuracy metrics	Supports career decision-making by evaluating influencing factors via ML
[51]	Placement prediction using academic and aptitude data	Support Vector Machine (SVM)	Helps institutions focus on high-potential students; improves placement rates
[52]	Academic risk detection and intervention	Random Forest (RF)	Identified students at risk early in the course, supporting timely academic support.

4.3 Selection Methods of Student Performance in Educational Data Mining

Preethi, J., & Maheswari, S. [53] suggested Educational organisations play an important part in a country's growth, yet forecasting student accomplishment is difficult due to the abundance of data in the educational database. Educational Data Mining Techniques can increase student performance and accomplishment by collecting information such as name, CGPA, and arrears. The Naïve Bayes method is used to analyse and classify student data by predicting suitable attributes. This strategy can be used with pupils who have divergent data on their previous and present academics. The suggested methodology employs a classification strategy to enhance estimation methods for forecasting placements, allowing for a link between academic success and placement achievement in campus selection. This strategy might help both instructors and academic institutions.

Walsh, S. & Cullinan, J. [54] examined the elements that affect an Irish student's decision to attend a higher education institution (HEI). It examines global research on these elements, emphasizing the effects of parents, siblings, and peers as well as aspects of domestic life. According to a discrete choice experiment, students place the most importance on work placement possibilities and course reputation. For HEIs, students often favor universities over technical institutions. The study finds that different students have different preferences for HEIs.

In education, data mining is essential for predicting student success and addressing deficiencies. Numerous elements, such as social, economic, personal, cultural, geographic, and institutional settings, might affect a student's academic success or dropout rate. Machine learning methods can help identify traits that affect a student's success. Comparing the results of machine learning classification models for forecasting student academic progress is the aim of this study. Results are compared utilizing fuzzy logic, hybrid, and classical logic approaches by Kaur, A. [55]. The main goal is to identify the most effective model for predicting student performance so that focused efforts may be made to improve performance and reduce dropout rates.

In order to forecast children's academic achievement, Yakubu, M. N., & Abubakar, A. M. [56] attempted to employ early detection factors such age, gender, high school exam results, region, and CGPA. To build a machine learning model, secondary data from a Nigerian institution was used. The results showed that age does not influence academic

achievement that kids in Years 3 and 4 are more likely to succeed academically, that female students are 1.2 times more likely to have a high CGPA, and that students with good JAMB scores are more likely to have a high CGPA.

Alam, A. et.al [57] discussed that Student success is critical at colleges and universities, and early detection of at-risk individuals can have a major influence on academic achievement. Machine learning approaches are becoming more popular for forecasting student achievement. However, these approaches are frequently restricted to computer science or artificial intelligence specialists. To adopt a successful data mining plan, educators must define student accomplishment, identify critical attributes, and select the right machine learning technique. This study attempts to give educators a complete approach on applying data mining techniques to predict student achievement. This program attempts to reduce the entrance barrier for data mining tools by analysing relevant literature and offering rationales for its selection.

Colleges are increasingly utilising predictive analytics to assist at-risk pupils. However, the majority of predictive analytics systems are private, with minimal transparency regarding their algorithms. Bird, K. A. et.al, [58] study analyses two aspects: sample and variable creation and predictive modelling methods. The results reveal that the relative ranking of students' anticipated college completion probability differs between models. While models trained on normal enrolment spells provide large increases, simple and sophisticated models perform similarly.

Machine learning is a branch of computer science that allows computers to learn on their own without the need for external programming. Kumar, V. U. *et.al*, [59] employs two approaches: supervised and unsupervised learning. Unsupervised learning use K-means and hierarchical clustering, whereas supervised learning employs Naive Bayes and Decision Trees. These algorithms are employed in many parts of student evaluation, including the recruiting process. In this scenario, K L University students' performance is assessed using these methods, highlighting the significance of machine learning in predicting output from inputs.

Data mining is the process of collecting useful information from massive data sets, allowing users to make educated decisions. K. Sreenivasa Rao and et al. [60] suggested using the To predict placement, use Education Data Mining (EDM) technologies. Higher education institutions utilize data mining (EDM) to forecast student achievement and

campus placement. Multiple Linear Regression, Binomial Logistic Regression, Recursive Partitioning, Regression Tree (rpart), Conditional Inference Tree (ctree), and Neural Network (nnet) are some of the machine learning algorithms that EDM employs in R Studio. Weka makes use of J48, Random Forest, Random Tree, and Naïve Bayes. Higher education institutions may give students better instruction by comparing the accuracy and performance outcomes.

Shoaib, M., et.al [61] discussed that Technology and data analysis have become indispensable tools for gathering, studying, and comparing student success levels in the classroom. Educational Data Mining (EDM) is a new trend in educational data mining approaches that aims to better understand student behaviour and the variables driving it. Learning patterns, culture, and teaching skills are all important components of effective EDM study. This research evaluates the usage of technology and data mining in EDM situations and their results. Previous study has been conducted to find the most effective way for analysing the learning environment and identifying characteristics that influence student academic achievement. Two cutting-edge models, decision tree and DBSCAN, are utilised to forecast educational institute performance with more precision.

A new machine learning model for forecasting undergraduate students' final test outcomes based on midterm exam scores was given by Yağcı, M. et al. [62]. The model uses a number of techniques, such as logistic regression, Naïve Bayes, support vector machines, random forests, k-nearest neighbors, and closest neighbor. 1854 students from a Turkish Language I course are included in the dataset. Using department, instructor, and midterm exam scores, the model achieved a 70-75% categorization accuracy. The development of a framework for learning analysis in higher education and the facilitation of decision-making processes depend heavily on this data-driven study. Additionally, it helps identify children who are at a greater risk of failing early on.

Dabhade, P. et.al [63] employed educational data mining to predict students' academic achievement at an Indian technical college. A dataset was collected using a questionnaire-based survey and an academic section. Data pre-processing and component analysis were used to eliminate anomalies and minimise data dimensionality. Python 3 was used to compare machine learning algorithms, and the support vector regression_linear approach outperformed others.

Palacios, C. A. et.al [64] forecasted student retention at various higher education levels

using machine learning methods. In every instance, the models—which include logistic regression, naive Bayes, random forest, decision trees, k-nearest neighbors, and support vector machines—achieve an accuracy of above 80%. Additionally, the models identify neighborhood poverty indexes and secondary school grades as important determinants. The findings apply to universities all throughout the world, including Chile, where dropout rates have an effect on productivity. Schools may take primitive measures and avoid dropouts by using student data to predict dropouts. In the example study, the algorithms perform better when the majority and minority classes are balanced.

Guleria, P. et.al [65] offered a career counselling model that integrates Explainable AI (XAI) and Machine Learning (ML) techniques. In order to help students get employment placements and make informed decisions regarding their professional growth, the framework looks at educational components. An educational dataset including academic and employability features is examined by ML-based White and Black Box models. With scores of 91.2% and 90.7%, respectively, the Naive Bayes model beat the Logistic Regression, Decision Tree, SVM, KNN, and Ensemble models in terms of Recall and F-Measure. The framework functions similarly to an expert system, giving decision help and recognising the challenges encountered by humans.

Educational institutions strive to give students with the greatest educational experience possible as well as excellent placement possibilities. **.** Ingale, Y et.al [66] identify students who require further assistance and taking necessary measures to improve their performance is critical for successful placement. Academic accomplishment and campus selection are complex concerns in the educational system. The suggested student prediction system is an important tool for differentiating student data based on performance. This system utilises machine learning techniques such as Naïve Baiyes, SVM, KNN, and decision tree to forecast student performance, identifying areas for improvement. This technology will benefit educational organisations by allowing them to better understand and enhance student performance**.**

Al-Zawqari et.al [67] presented an adaptable framework for forecasting students' academic achievement in online and mixed learning. The approach relies solely on raw data, with no feature engineering, to ensure model interpretability. The model is tested on the Open University learning analytics dataset (OULAD) with random forest and artificial neural networks. The results suggest that the flexible feature selection framework outperforms conventional techniques based on manual feature engineering. The model has excellent

accuracy rates for children at danger of failing and dropping out, but it performs poorly in predicting high-achieving pupils.

According to Albreiki, B. et.al [68] Educational Data Mining (EDM) is critical for enhancing the learning environment by employing innovative approaches and techniques. This comprehensive evaluation of EDM literature from 2009 to 2021 demonstrates that several Machine Learning (ML) approaches are employed to predict student dropouts and at-risk pupils. The majority of research employ data from student databases and online learning systems. ML approaches have been shown to be critical in anticipating these difficulties and boosting student performance, thus improving the overall quality of education in modern academic institutions.

Khan, A., et.al [69] conducted a comprehensive evaluation of educational data mining (EDM) studies on student performance in classroom settings. The review is on identifying predictors, prediction methodologies, timing, and prediction goals. It is the first systematic review of EDM research that takes into account classroom learning and the time dimension. The meta-analysis shows that researchers obtain substantial prediction efficiency over the course duration, while performance prediction before to course start requires particular attention. The report conducts a review of 140 papers in this field, emphasising the need of understanding the temporal element of prediction in EDM.

The study by Bhargavi, M., [70] employed machine language (ML) algorithms to forecast student placement results, an important consideration for educational institutions in today's competitive employment market. The decision tree model has the best accuracy (83.72%), followed by Naive Bayes (81.40%). KNN and RF reached 79.07% accuracy, whereas ensemble approaches and gradient boosting got similar values of 81.40%. These machine learning tools can help enhance placement processes, identify at-risk pupils, and adjust treatments.

Educational data characteristics are frequently unclear, which leads to noise and dimensionality issues. To overcome these concerns, Abdullahi, W [71] created a reliable feature selection technique called Cascade Bi-Level (CBL). This method combines Particle Swarm Optimization (PSO) on the second level with the Relief technique on the first level. The UCI student performance dataset was used to test the CBL technique, which outperformed the single-level strategy by 94.94%. For binary classification tests, this performs better than the 93.67% accuracy of the single-level PSO. By properly predicting

student achievement, the CBL method may lessen the limitations of single-level systems. This approach performs better than previous methods that disregard classifier interaction and feature dependency.

Tarekegn G.B. et al. [72] proposed an approach with the help of data mining classification techniques to predict the chance of students to get the department of their preference after the entrance exam result. The approaches using the algorithm for prediction are Random Forest algorithm , J48 and Naïve Bayes algorithm to build model , from which the algorithms which is suggested to be best is the random The result shows that many students get their first preference and few of them were not given their first preference.

Lulla Chandini et al. [73] proposed the prediction on academic performance through data mining and machine learning. Author proposed this system is basically designed for the B.Tech final year students of Sinhgad Academy of Engineering (S.A.E), where they collected student dataset from last 3 to 5 years and perform the prediction with the ID3 technique of data mining with the attributes like B.Tech aggregate, attendance, co - curricular activities, family background, backlogs, AMCAT results etc. The main objective is to analyze the overall performance of the student and how to improve the individual performance in which they fell behind. In this the results were classified in four categories i.e. Poor, Average, Good and Excellent.

Shah M. B. et al. [74] presented an approach which makes predictions to find out the student performance including various factors like interest, opinion etc. The approach uses machine learning and deep learning to find out the best result to predict student accomplishment like decision tree, SVM, Logistic regression, XG-Boosting, Gradient Boosting, Random Forest among which Gradient Boosting provides better results.

For a number of reasons, such as a lack of resources, apathy, gender issues, and a shortage of quality schools and teachers, many Indian children drop out of school. It is increasingly challenging to select the right job path as the number of employment alternatives rises. 40% of students feel unsure about their job choices, which might result in the wrong pathways and unexpected consequences, according to a poll conducted by the Council of Scientific and Industrial Research (CSIR). Sah, U. K., & Singh, A. used the Adaboost, Random Forest, Support Vector Machine, and Decision Tree algorithms together with the Python programming language. [75] suggested approach applies machine learning principles to forecast students' future educational attainment.

Table 4.2 Comparative Summary of Literature in Predictive Analytics in Educational Data Mining

Ref	Focus Area	Methods/Techniques Used	Key Findings/Contributions
[53]	Predicting student placement using educational data	Educational Data Mining, Naïve Bayes Classifier	Classification of student using academic data (e.g., CGPA, arrears) for predicting placement; supports educational planning.
[54]	Factors influencing students' choice of higher education institutions in Ireland	Discrete Choice Experiment	Work placement opportunities and course reputation are top factors; university preference over technology institutes noted.
[55]	Forecasting student achievement and dropout risk	Classification Models (Traditional, Hybrid, Fuzzy Logic)	Comparative analysis to identify the best risk and performance predictors.
[56]	Predicting academic success using demographic and academic indicators	Machine Learning Classification Model	Found gender and test scores predictive; age not significant.
[57]	Application of ML for early identification of at-risk students	Data Mining + ML (approach-driven rather than technique-specific)	Educator-friendly guide for implementing data mining in academics.
[58]	Evaluation of predictive models for college completion	Simple and Sophisticated ML Models	Explained model variations and transparency in institutional predictive analytics.

[59]	Student performance evaluation and recruitment	NaiveBayes, Decision Tree,K-means, Hierarchical Clustering	ML algorithms effectively predict student performance at KL University
[60]	Forecasting student performance and placement	J48, Naïve Bayes, Random Forest, Random Tree (Weka); Multiple Linear Regression, Logistic Regression, etc. (R)	Compared models in Weka and R Studio for performance and accuracy
[61]	Educational environment analysis and institutional performance	Decision Tree, DBSCAN	Decision Tree and DBSCAN used to predict institutional performance accurately
[62]	Predicting final test scores from midterm results	Random Forest, KNN, SVM, Logistic Regression, Naive Bayes	Achieved 70–75% accuracy using academic and demographic data (n = 1854)
[63]	Academic achievement prediction in Indian technical college	Support Vector Regression (Linear)	SVR (Linear) outperformed other models in predicting student academic performance
[64]	Student retention prediction	Decision Tree, KNN, Logistic Regression, Naive Bayes, Random Forest, SVM	Models achieved >80% accuracy; key predictors: educational scores, poverty index
[65]	Career counselling	Naive Bayes, Logistic	Naive Bayes achieved

	and placement	Regression, Decision Tree, SVM, KNN, Ensemble Models (with XAI)	highest Recall (91.2%) and F-Measure (90.7%)
[66]	Academic Database Prediction	Naïve Bayes, SVM, KNN, Decision Tree	Helps identify weak students and areas for improvement for placement preparation
[67]	Academic Performance in e-learning	Random Forest, Artificial Neural Networks	Adaptable framework effective without manual feature selection
[68]	Prediction on Academic Outcome	ML techniques across multiple studies	Review of EDM studies (2009–2021); ML key in predicting dropout and performance
[69]	Prediction on Student Performance	Varies across 140 studies	First EDM review considering prediction timing; stresses early prediction complexity
[70]	Analysis on academic performance and recruitment	Decision Tree (83.72%), Naïve Bayes (81.40%), KNN, RF, Ensemble, Gradient Boosting	ML models can improve placement forecasting and early risk identification
[71]	Cascade feature selection technique in predicting student performance	Not explicitly mentioned, binary classifier implied	CBL outperformed single-level PSO, addressing feature interdependence and classifier interaction
[72]	Forecasting student placement into academic departments.	Decision Tree, Naïve Bayes	Departmental placements based on students' academic attributes.

[73]	Prediction of academic performance for final-year B.Tech students	ID3 (Decision Tree) using data mining	Classified students into four categories (Poor, Average, Good, Excellent) using parameters such as aggregate score, attendance, co-curricular, backlogs, AMCAT results, and family background; aimed to improve individual student performance
[74]	Student performance prediction based on academic and psychological factors	Decision Tree, SVM, Logistic Regression, XGBoost, Gradient Boosting, Random Forest	Gradient Boosting outperformed other models; considered factors like student interest and opinion for more holistic performance prediction
[75]	Career path prediction for students	Decision Tree, RF, SVM, Adaboost (Python)	Addresses issues like career confusion; uses ML to forecast educational attainment

4.4 Research Gaps

There are still a number of research gaps despite the fact that machine learning and feature selection techniques have been used in numerous studies to forecast student performance and placement outcomes. Since many traditional models rely on single-level feature selection techniques, which can lead to decreased prediction accuracy, a major limitation is the insufficient consideration of feature interdependence and classifier interaction. And, a large portion of the current research relies on manual feature engineering, which restricts scalability and adaptability and introduces subjectivity, particularly in dynamic academic

environments.

Although they show promise, studies that do not use manual feature selection—for example, those that use raw data for model interpretability—still have difficulty correctly predicting high-achieving students. Furthermore, the majority of current works do not fully take into account both quantitative and qualitative characteristics that are essential for placement prediction, such as creativity, research potential, and extracurricular involvement.

As demonstrated by localized surveys such as the one carried out in Chhattisgarh, there is also a need for additional region-specific research that tackles the difficulties of data collection. Finally, despite evidence that suggests the timing of predictions has a significant impact on their accuracy, the temporal aspect of prediction—particularly in classroom-based settings—remains understudied.

For accurate student performance and placement forecasting, these gaps underscore the need for more resilient, flexible, and context-sensitive models.

CHAPTER – 5

PROPOSED METHODOLOGY

5.1 Introduction

This chapter discusses how to create a data-driven hiring framework that enhances the traditional campus placement process. The study used machine learning algorithms to forecast student placement results and is based on an empirical and scientific approach. The technique encompasses data collection, pre-processing, feature selection, model training and validation, and performance evaluation.

5.2 Research Design

The study employs applied machine learning in tandem with a quantitative research approach. Its primary goal is to create predictive models that assist in determining students' selection possibilities based on their academic, assessments, achievements and many other criteria which can fulfil the company requirements. The primary objective is to employ intelligent, scalable, and ethical technologies to reduce recruiting bias and improve placement efficacy.

5.3 Data Collection

The fundamentals of this framework requires gathering the high quality datasets from the universities and the companies. Data set collected from some recruiters on a confidential basis, as well as some data from Kaggle. These data should contains:

- Academic records (e.g., CGPA and semester grades)
- Technical skills and certifications
- Work experience (if applicable)
- Extracurricular Activities and Achievements

- Company hiring patterns

The primary data items employed in the study are:

Data obtained from companies- Parameters for evaluation include skill requirements and selection criteria.

Student data: Data has been gathered from some companies to know their academic performance, technical and soft skills, extracurricular activities, certificates, and evaluation results.

Evaluation data: Student scores in company-conducted placement tests.

At first number of parameters that should be included in the placement process is decided and according to that maximum possibilities is generated in the binary form (i.e. 0 & 1), Then the data is taken from the student and company which is then cleaned and formatted to manage missing values, remove extraneous objects, and convert category variables to numerical or binary forms for improved model compatibility.

5.4 Feature Engineering and Selection

Relevant features were chosen based on statistical significance and conformity to industry hiring standards in order to maximise model performance. Recursive feature elimination and correlation analysis were among the methods employed. The model is trained using the greatest number of possible outcomes for the dataset using a specific number of parameters in the form of 0 and 1.

Psuedocode for maximum possibilities generation:

```
Input: num_parameters  =  n total_combinations = 2^num_parameters
combinations = [] for i in range(total_combinations):
binary_string = format(i, '0{}b'.format(num_parameters))
combination = [int(bit) for bit in binary_string]
combinations.append(combination)
output_file = 'combinations.csv'
with open(output_file, 'w', newline='') as file:
 csv_writer = csv.writer(file)
```

```
csv_writer.writerows(combinations)
print("All generated combinations have been saved to", output_file)
```

The characteristics are chosen based on the company's requirements which are prioritized as per the need, such as academic marks, undergraduate scores, HR interview scores, or any other feature as needed. The objective is to minimize dimensionality while retaining influential factors for training. Binary variables (e.g., 0 for "unselected" and 1 for "selected") maintain consistency among algorithms.

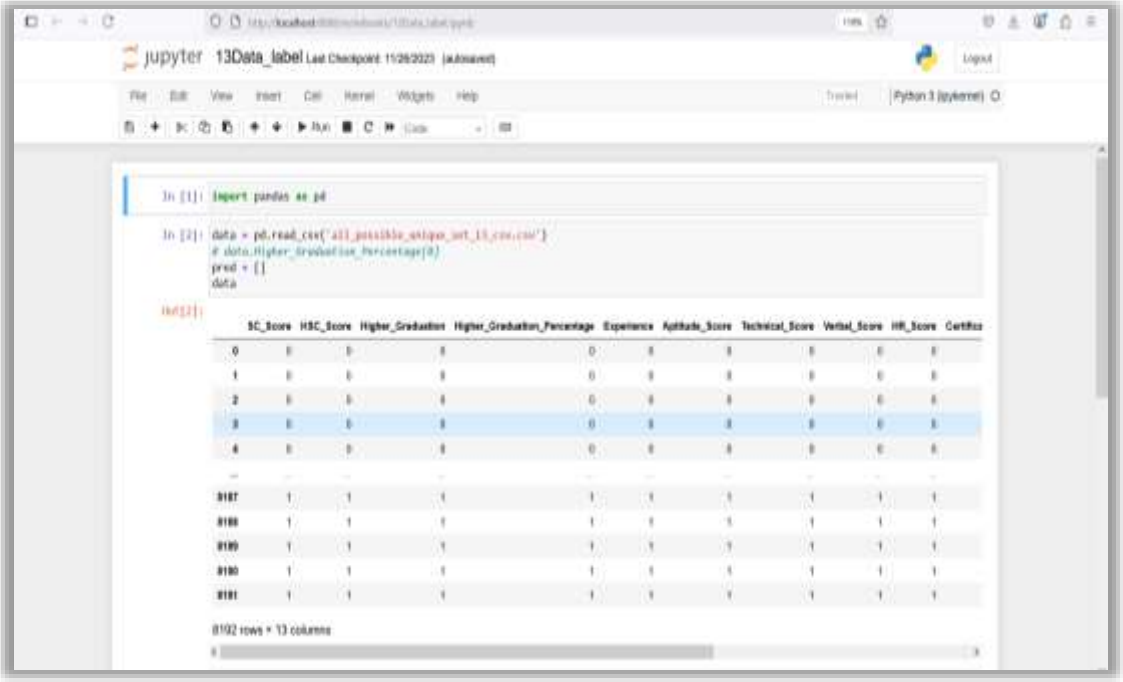

Figure 5.1 Maximum Possibilities in 0 and 1 format

5.5 Data Generation for Model Training

The generated dataset was split into training (80%) and testing (20%) sets to guarantee reliable model training. To replicate actual recruiting circumstances, a range of recruitment scenarios were simulated. The models' generalizability was enhanced and class imbalance was lessened by the balanced dataset.

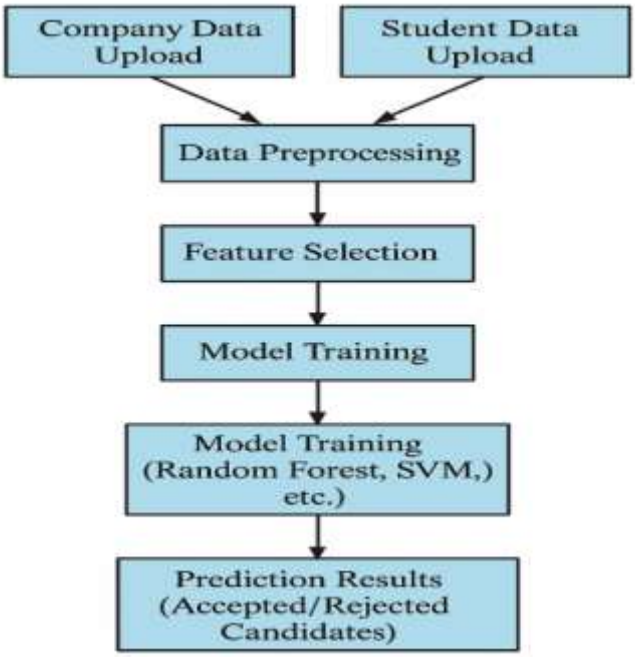

Figure 5.2 Concept map

Making predictions means constructing a large dataset that includes a number of possible outcomes based on the supplied factors.

Prediction algorithms are successfully trained and tested using this dataset. The purpose of these models is to calculate the probability of choosing a candidate based on the given criteria. The process starts when business data, including the organization's established assessment criteria, is uploaded. Additionally, evaluation and student data—which include assessment and learning data—are supplied. A prediction result that clearly classifies candidates as rejected or selected is the process's output.

5.6 Evaluation Metrics

A variety of evaluation metrics are used to gauge the efficiency and dependability of the machine learning models used for candidate prediction. These metrics offer a thorough grasp of the models' performance in terms of classification accuracy and generalizability to new data.

Accuracy: Accuracy, which is the ratio of correctly anticipated events (including true positives and true negatives) to the total number of forecasts, indicates the model's overall soundness. Even though it provides a broad idea of performance, unbalanced datasets may

be deceptive.

Precision: The precision metric quantifies the percentage of accurate positive predictions among all the positive predictions generated by the model. The issue, "Of all the candidates predicted as selected, how many were actually selected?" is addressed. Low false positive rates are correlated with high accuracy.

Recall (Sensitivity): Recall measures the model's capacity to detect real positive cases. The real positives to genuine positives ratio is what it is. It represents the proportion of candidates with genuine qualifications that the model found during the hiring process.

F1-Score: The harmonic mean of accuracy and recall is known as the F1-score. When there is an unequal distribution of classes or when there are a lot of false positives and false negatives, it offers a fair metric. It is particularly helpful in situations when choosing an unqualified candidate (false positive) is more expensive than missing a qualified one (false negative).

Region below the AUC (ROC Curve): To plot the true positive rate versus the false positive rate, the Receiver Operating Characteristic (ROC) curve is used at various threshold levels. The AUC indicates the model's capacity to distinguish between classes. While a score around 0.5 suggests negligible discriminative capacity, an AUC value near 1 indicates excellent model performance.

Confusion Matrix: A tabular summary that provides a more detailed view of classification results is called a confusion matrix. For every algorithm, it shows the numbers of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). Based on the particular needs of the application, this breakdown assists in determining the kinds of errors the model is producing and in choosing the best model for deployment.

A thorough analysis of each algorithm's performance is carried out by combining these evaluation metrics, guaranteeing that the chosen model not only maintains a balanced performance across all important classification aspects but also achieves high accuracy.

5.7 Implementation

During the implementation phase, the established prediction framework is applied to real-world data acquired from corporations and academic organisations. This integrated dataset

includes student academic and evaluation data as well as specific recruiting conditions specified by corporations. The chosen machine learning models are then applied to this real-time data after being trained on a variety of data sets.

After training, the models are evaluated against experimental or unknown data to determine their prediction potential in real-world situations. The purpose is to simulate a real-world recruiting scenario so that the system may assist with making educated applicant selection judgements.

FUNCTIONAL REQUIREMENTS

- ✓ Interface
- ✓ Providing company specifications
- ✓ Feeding Student Data into the System
- ✓ Uploading the company-conducted evaluation data
- ✓ Model prediction

Interface:

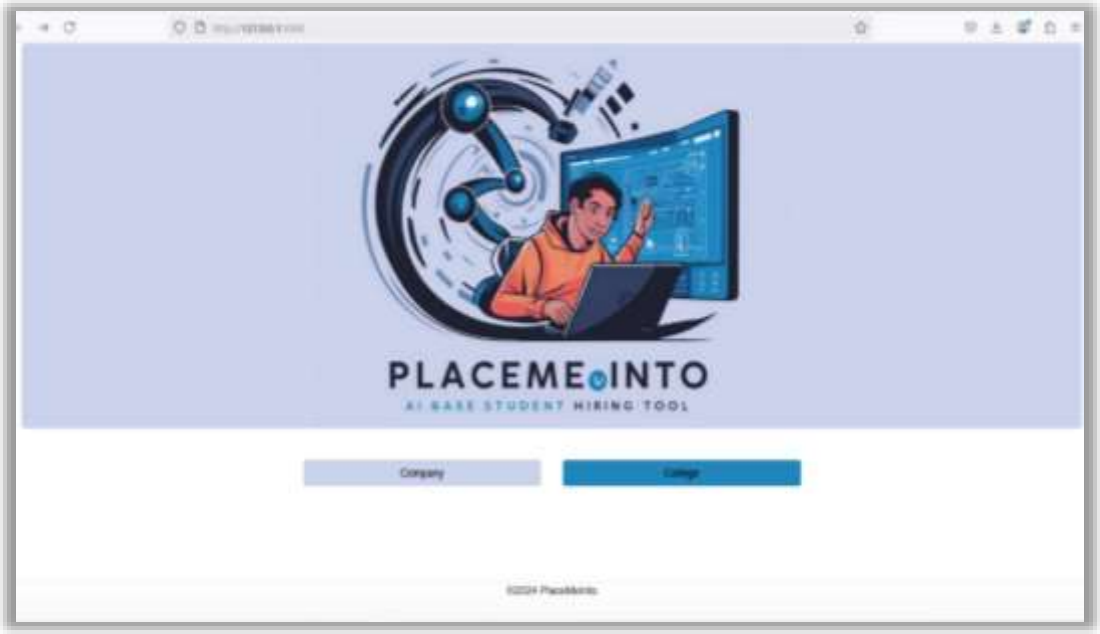

Figure 5.3 Interface

Providing company specifications:

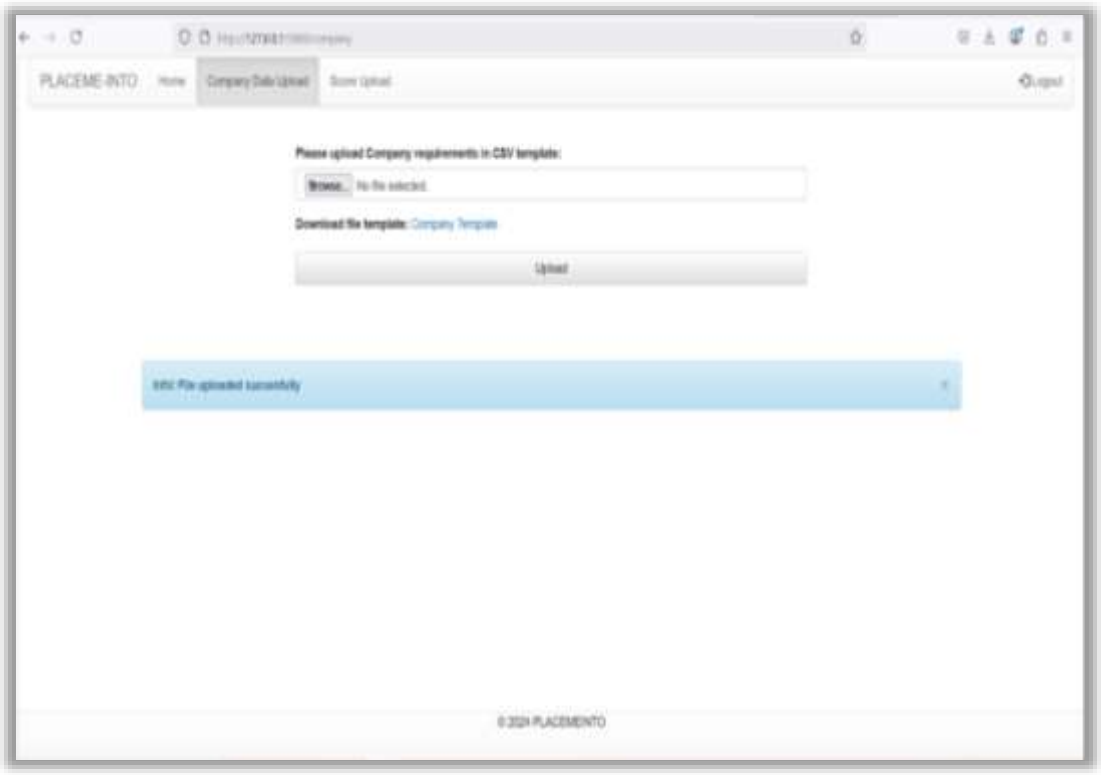

Figure 5.4 Company Criteria Upload Screen

Feeding Student Data into the System:

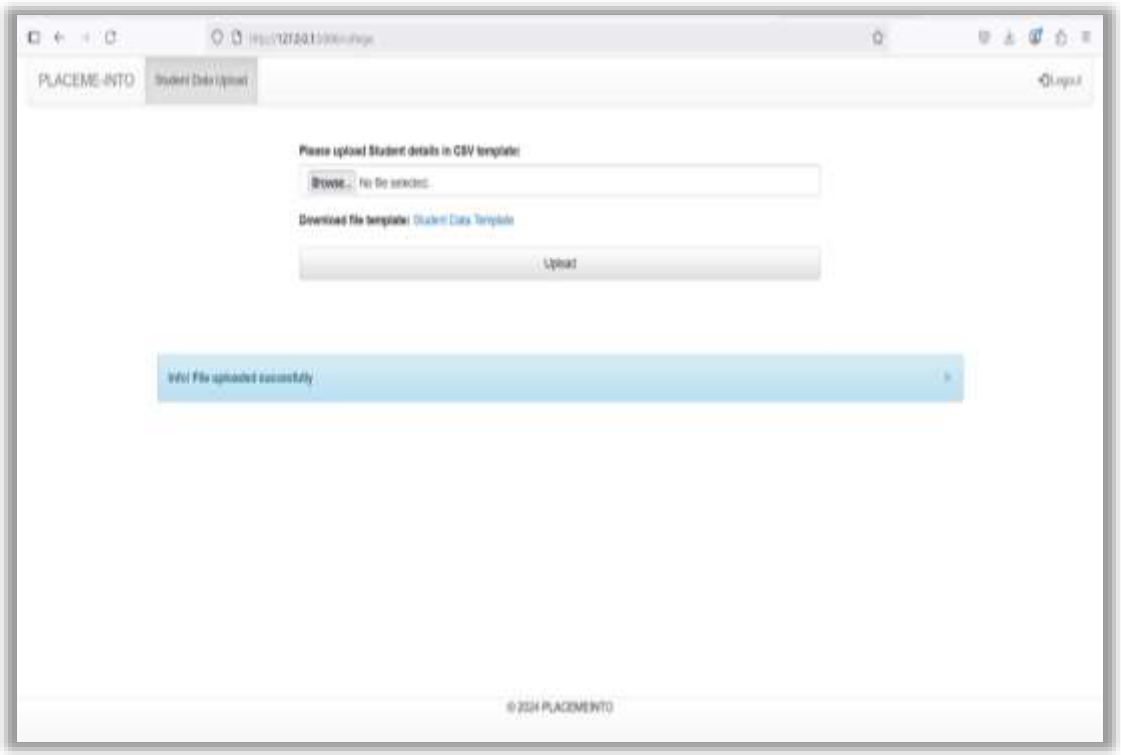

Figure 5.5 Student Dataset Upload Screen

Uploading the company-conducted evaluation data:

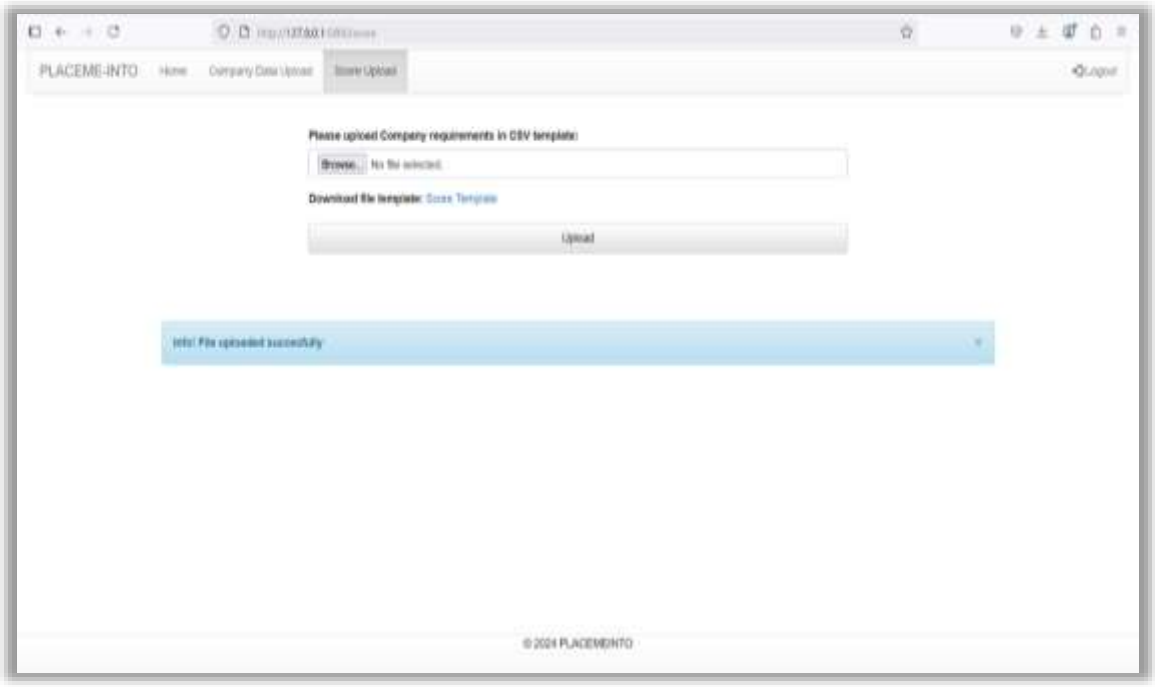

Figure 5.6 Company Evaluation Score Upload Screen

Model prediction:

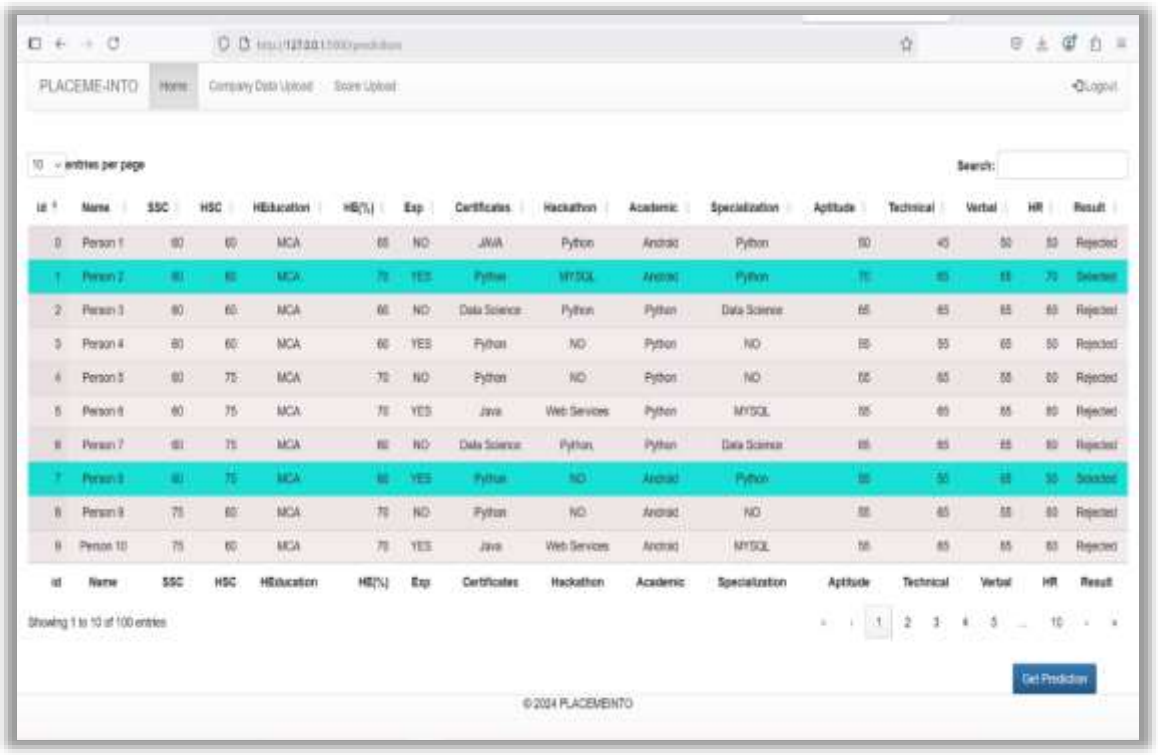

Figure5.7 Output of prediction

5.8 Applied Techniques

Some of the machine learning techniques were implemented to predict student placement outcomes. The models used in the study include:

- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- Boosting Algorithms (e.g., XGBoost, AdaBoost)
- Logistic Regression
- KNN

5.8.1 Decision Tree Classifier

A decision tree is a type of supervised machine learning approach that models decisions, outcomes, and predictions using a flowchart-style tree structure. An algorithmic procedure (series of if-else statements) is used to create such a tree by identifying ways to segment, categorise, and visualise a dataset based on various circumstances.

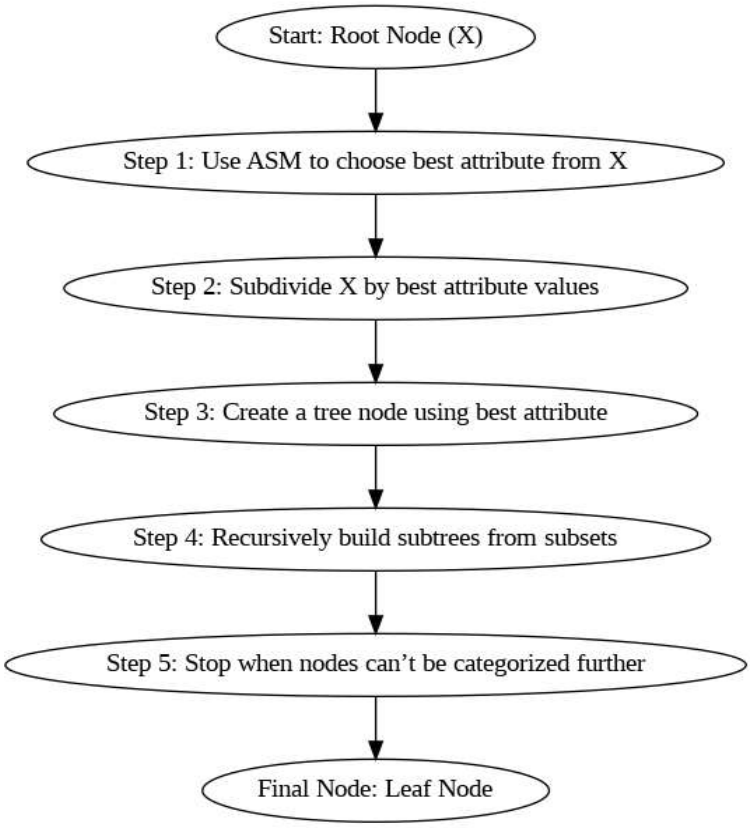

Figure 5.8 Decision tree flow

5.8.2 Random forest

A popular supervised machine learning method for solving regression and classification issues is random forests.

Random forest approaches are a form of classifier that uses a bootstrap sample and decision trees to solve regression or classification issues. A third of the training sample is reserved for testing and is known as the out-of-bag (oob) sample. By adding randomization, feature bagging improves dataset diversity and reduces decision tree correlation. The class prediction is determined by the nature of the problem. While individual decision trees are averaged for regression tasks, a majority vote decides the projected class for classification tasks. The out-of-bag data is then used for cross-validation to generate the final forecast.

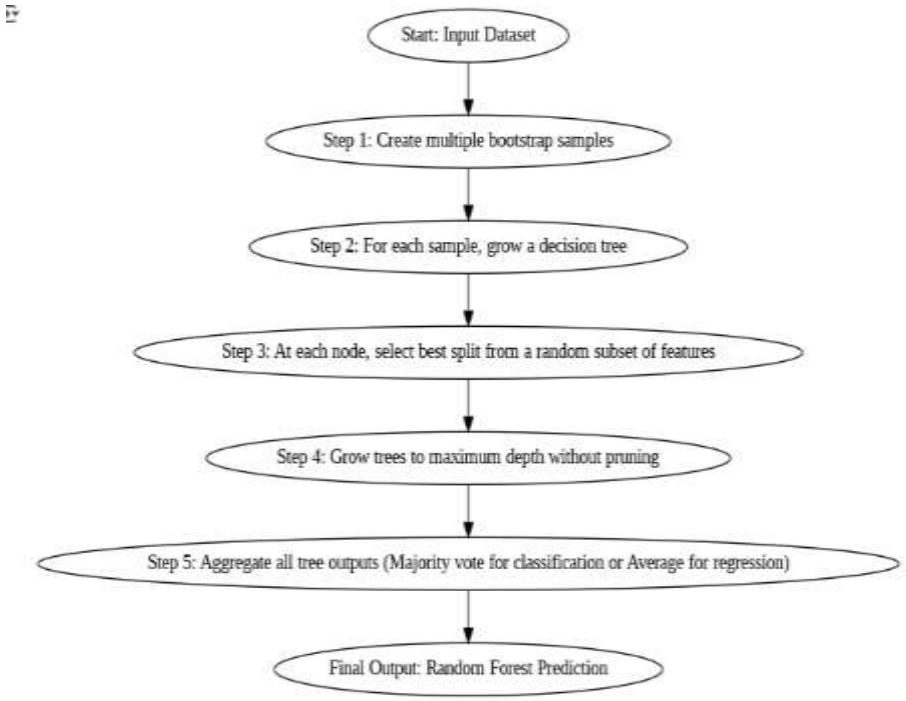

Figure 5.9 Random forest flow

5.8.3 SVM

This is a technique that classifies data by finding the best line or hyperplane in an N-dimensional space that optimizes the distance between each class.

SVMs are often employed in classification issues to choose the best hyperplane that

optimizes the margin between the nearest data points of opposing classes. Depending on the number of features in the input data, the hyperplane may be a line in two dimensions or a plane in n dimensions. This enables the algorithm to determine the optimal decision border between classes, enhancing its capacity to provide accurate classification predictions and generalize to new data. Lines that cross the data points with the biggest margin and run parallel to the ideal hyperplane are known as support vectors. This enhances the algorithm's capacity to perform accurately on previously unseen inputs and delivers reliable and precise class labels. by enabling it to determine the optimal decision boundary between classes. Lines that run parallel to the ideal hyperplane and span the data points that make up the biggest margin are known as support vectors.

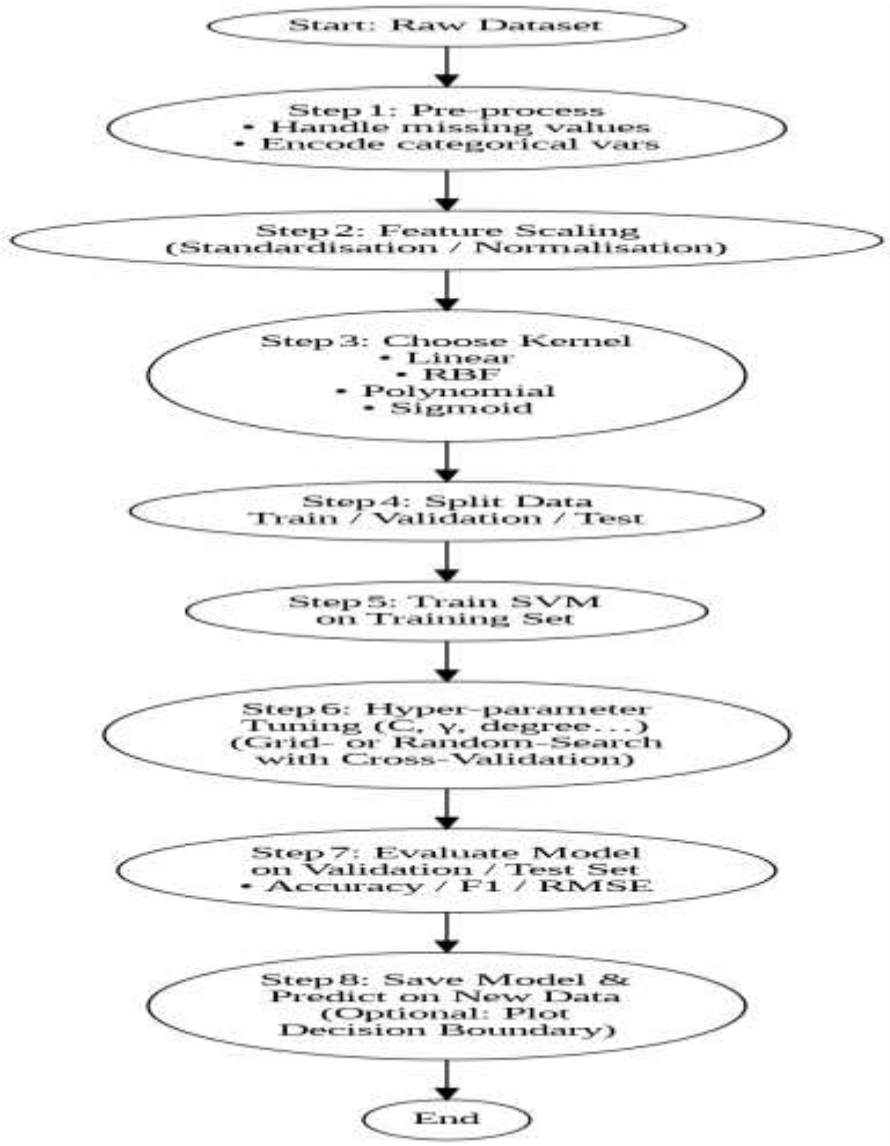

Figure 5.10 SVM Algorithm flow

When data cannot be separated linearly, the SVM approach employs kernel functions to transform the data into a higher-dimensional space for linear separation, even though it can manage linear or nonlinear categorization issues. The kernel function used is determined by the use case and the properties of the data.

5.8.4 Boosting Algorithms (e.g., XGBoost, AdaBoost)

XGBoost is an open-source machine learning toolkit that utilizes gradient-boosted decision trees, a supervised learning approach based on gradient descent, to improve efficiency, speed, and scalability in big datasets. It handles missing values in sparse data using a sparsity-aware technique, which assigns data points to the default direction when a value is absent. XGBoost is based on the DMatrix data structure, which has been tuned for training speed and memory economy.

The use case determines the model's target function, which might be binary or multi-class classification.

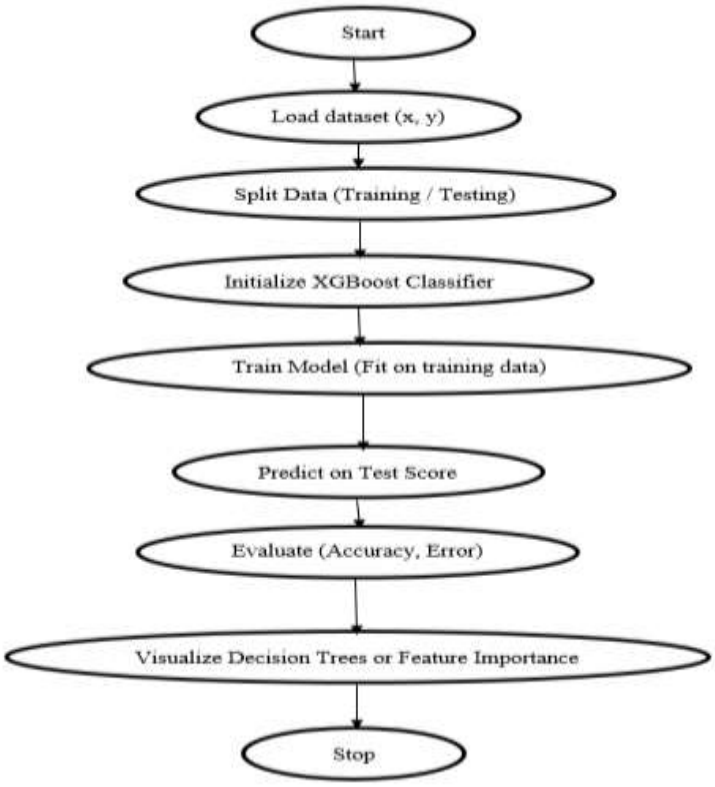

Figure 5.11 XGBoost flow

5.8.5 Logistic Regression

When the dependent variable is dichotomous (binary), logistic regression is the most effective regression technique to use. Like other regression analyses, it is a predictive analysis. It is employed to describe data and elucidate the relationship between one or more dependent binary variables and one or more nominal, ordinal, interval, or ratio-level independent variables.

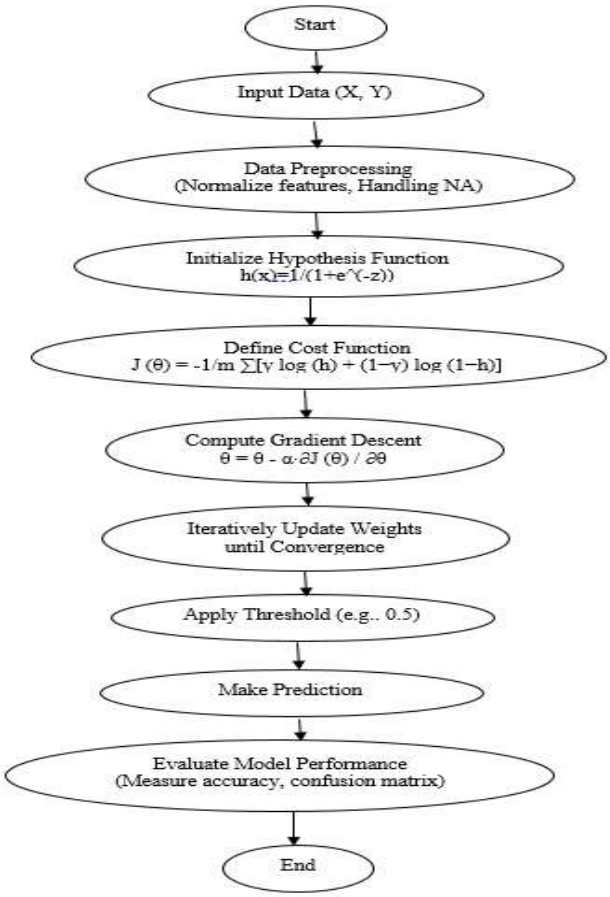

Figure 5.12 Logistic Algorithm flow

5.8.6 K NN

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm that may be used for classification as well as regression. It identifies which "k" data points are the most comparable to a certain input. KNN predicts outcomes by either selecting the most frequent class (for classification tasks) or calculating the mean value (for regression tasks). As a non-parametric, instance-based learning approach, it does not rely on any assumptions regarding the data's underlying distribution.

Choice of K

By experimenting with a few numbers, the optimal k value—the number of nearest neighbors taken into account—is found to produce the most accurate forecasts with the fewest errors. While large k values are noisy and improve prediction accuracy by providing more data to compute modes or means, low k values cause instability in predictions because of high variation, complexity, and low bias. However, if the k value is too large, the model may not be sophisticated enough to fit the training data effectively, which could lead to low variance, low complexity, and high bias.

Figure 5.13 KNN Algorithm flow

5.9 Model Validation and Performance Evaluation

To guarantee the dependability and utility of the suggested placement prediction system, a thorough model validation and performance evaluation procedure was conducted. Using common classification metrics including accuracy, precision, recall (sensitivity), and the F1 score, each machine learning model's performance was assessed. While accuracy represents the percentage of all correct predictions produced by the model, precision is the number of predicted positive cases that were correct. The F1-score provides a single statistic for evaluating the overall performance of the model by striking a balance between accuracy and recall. The model's recall gauges its capacity to identify all pertinent positive cases.

To evaluate model performance, we have used range of statistical metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5.10 Proposed Methodology

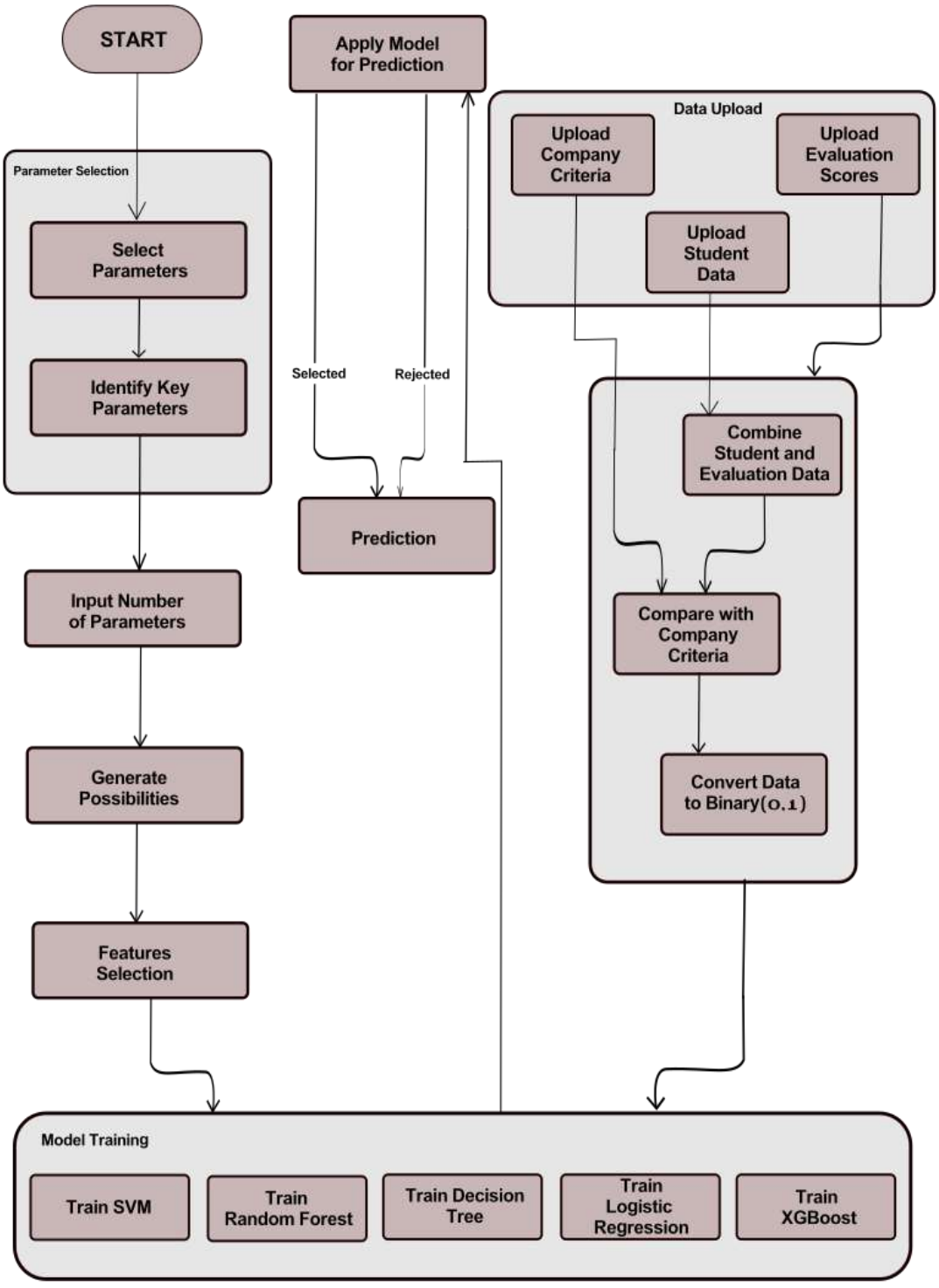

Figure 5.14 Methodology of the study

5.11 Comparative Analysis: Existing vs. Proposed Methodology

A comparative study was carried out to highlight the advantages of the proposed framework over standard placement strategies.

Table 5.1 Existing vs. Proposed Methodology

Aspect	Existing Methodology	Proposed Methodology
Candidate Evaluation	Generic aptitude and technical tests	Tailored, parameter-driven assessments
Focus Area	Primarily technical knowledge	Holistic (technical, soft skills, cultural fit)
Bias	Subjective and prone to favoritism	Objective, data-driven decision-making
Industry Alignment	Often misaligned with dynamic industry needs	Continuously aligned with employer expectations
Scalability	Limited adaptability across job roles	Easily scalable across various industries
Time Efficiency	Time-consuming and repetitive	Streamlined with predefined screening parameters
Inclusivity	Favours urban, well-connected students	Focuses on potential, diversifying the talent pool
Training Costs	High due to mismatched skills	Lower due to better student-role fit
Student Preparation	Ambiguous expectations	Transparent, goal-oriented preparation parameters
Retention and Performance	High attrition due to poor job-role fit	Increased retention and motivation through skill-role alignment

5.12 Summary

To improve campus placement practices, this chapter discusses a data-driven hiring architecture that employs machine learning methods. The platform makes advantage of high-quality information from businesses and universities, including job descriptions, technical ability, work experience, and academic records. It evaluates accuracy, recall, and AUC and is trained with a specific set of parameters. To forecast student recruitment results, samples were used with machine learning methods like Random Forest Classifier, Support Vector Machine (SVM), Boosting approaches, Logistic Regression, and Decision Tree Classifier.

CHAPTER – 6

RESULT ANALYSIS

6.1 Introduction

This chapter outlines the results of this study's machine-learning method for forecasting campus placement recruiting and selection results. We evaluated each model's capacity to predict which applicants would receive job offers after building and training many classifiers using carefully selected employer and student data. The chapter illustrates how the suggested approach improves hiring equity, more closely complies with industry standards, and expedites decision-making for recruiters and campus placement managers by contrasting these indicators with baseline placement numbers.

Campus placement, also known as campus recruiting, is a procedure in which businesses visit college campuses to interview candidates and make employment decisions. These placements can be held on campus, with prominent firms and official partners interviewing prospective individuals and evaluating their potential. Firms give employment offers following rounds of interviews and assessments, outlining explicit work requirements, remuneration scale, and other perks.

On-campus placements occur in a college or organisation, where applicants are responsible for finding work after finishing their degree. Pool-campus placements are organized by various educational institutions, with employers performing interviews and evaluations with students from participating universities. This enables businesses to benefit from a larger talent pool, particularly smaller or newly established businesses seeking roles with a diverse set of skills and capabilities.

College campus recruitment programs seek to benefit both employers and students by providing employment opportunities, corresponding students with satisfactory positions, exposing them to various vocations, educating graduates, intensifying the hiring process, assisting with workforce services, promoting professional connections, and filling knowledge gaps. However, some students may not be able to secure campus placement owing to weak academic achievement or technical competence. To guarantee that students

with strong talents do not fall behind, recruiters should evaluate all factors, including academic achievement, technical skills, examinations, and interviews.

6.2 Data

Recruiters actively engaged in the hiring process in Ahmedabad and the surrounding areas provided us with sensitive information. Furthermore, open platforms like Kaggle provided some of the data. An algorithm was created to produce all possible scenarios of the given dataset as per the number of parameters taken by the company for the selection of the student which includes academic, achievement and other test score evaluated by the recruiters in the form of binary as per the criteria given by the company for the recruitment of the student through campus placement.

6.3 Algorithm development

The data pre-processing step is when the machine learning method for recruitment and selection prediction in campus placement begins. In this scenario, raw datasets containing student academic and technical information, as well as job requirement attributes, are integrated using common identifiers such as department or academic year. Null values in the combined dataset are processed by filling in the mode or assigning an "unknown" category to categorical columns, while numerical columns are imputed with mean or median values. Numerical variables are then scaled or normalized to ensure consistency, and categorical variables are encoded using appropriate methods such as ordinal or one-hot encoding. Training and testing sets are then generated from the preprocessed data which is approx 8000.

After preprocessing, a series of machine learning techniques, like XGBoost, Random Forest, SVM, Decision Tree, KNN and Logistic Regression, are trained using the training data in the model training phase. Every model is saved for performance comparison after being fitted to the dataset. Using important performance indicators like accuracy, precision, recall, and F1-score, the model evaluation phase evaluates each trained model's predictive ability on the test data. The model that performs the best is chosen for additional optimisation based on evaluation.

The chosen model is fine-tuned using methods such as Grid Search or Random Search with cross-validation to enhance its predictive performance during the hyper parameter tuning

stage. After the best model has been determined, it moves on to the phase of deployment and additional analysis. Here, a decision support system incorporates the trained model to predict student placement outcomes in real time. Further analysis of the outcomes and forecasts can also reveal important placement-driving characteristics. A strong and data-driven strategy for forecasting recruitment and selection results during campus placements is ensured by this structured pipeline.

At a fictitious campus, the dataset provides a thorough description of student profiles and placement results. It is intended to examine a number of variables that affect campus placement, such as specialization, work experience, academic standing, and compensation offers. Such a dataset can be used for career counselling programs and predictive modelling, and it is essential for comprehending patterns and predictors of successful student placement.

The dataset, includes a wide range of academic, professional, and demographic characteristics. It contains the percentages for Classes 10 and 12 (ssc_p, hsc_p), the corresponding school boards (ssc_b, hsc_b), and the higher-secondary stream (hsc_s). The degree percentage (degree_p) and major category (degree_t) provide information about undergraduate performance, and workex flags as previous work experience. The test score (etest_p) reflects employability aptitude. The dataset records specialization (specialization) and percentage (master_p) at the postgraduate level and some other parameters which are needed by the company. Importantly, it documents the status of the placement and, for those who are placed. When taken as a whole, these variables provide a condensed but thorough picture of the variables that influence campus placement success.

6.3.1 Decision Tree Classification

Table 6.1 Decision tree Classification

Decision Tree Classification

Splits	n(Train)	n(Test)	Test Accuracy
39	6554	1638	1.000

Figure 6.1 Data split

The section of the dataset that was used to teach the model is represented by the blue portion of the bar. The bulk of the data, 6,554 observations, are included.

The data used to test or assess the trained model is displayed in this red section of the bar. There are 1,638 observations in all:

1638 (test) + 6554 (train) = 8192 (total)

Table 6.2 Confusion matrix

Confusion Matrix

		Predicted	
		0	**1**
Observed	0	0.98	0
	1	0	0.02

Table 6.2 suggests that

- 98% of class 0 were correctly predicted (true negatives).
- 2% of class 1 were correctly predicted (true positives).
- No misclassifications occurred.

Table 6.3 Class Proportion

Class Proportions

	Data Set	**Training Set**	**Test Set**
0	0.969	0.967	0.976
1	0.031	0.033	0.024

Table 6.3 demonstrates a strong class imbalance, with about 3% of all data in class 1 and 97% in class 0. The training set (96.7% vs. 3.3%) and test set (97.6% vs. 2.4%) both

preserve this skew almost identically, indicating that the data split was stratified and hence objective. Because class 0 dominance implies that a classifier can achieve high overall accuracy simply by predicting the majority class, it is recommended to take additional steps, such as resampling, class-weighted algorithms, or metrics focused on the minority class, to ensure the model learns to recognize class 1 cases.

Table 6.4 Evaluation matrix

Evaluation Metrics	0	1	Average / Total
Support	1599	39	1638
Accuracy	1.000	1.000	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.000
False Discovery Rate	0.000	0.000	0.000
F1 Score	1.000	1.000	1.000
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	1.000	1.000	1.000
Negative Predictive Value	1.000	1.000	1.000
True Negative Rate	1.000	1.000	1.000
False Negative Rate	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000
Threat Score	∞	∞	∞
Statistical Parity	0.976	0.024	1.000

A perfect evaluation profile is displayed in Table 6.4 All error measures (false-positive, false-negative, false-discovery, and false-omission rates) are zero, and all metrics (accuracy, precision, recall, F1-score, AUC, negative-predictive value, and true-negative rate) register at 1.000 for both classes. The model correctly classified every single case in the test set, which contained 1,599 instances of class 0 and only 39 instances of class 1. This resulted in infinite threat scores and rendered the Matthews correlation coefficient

undefined (NaN) because there were no errors to balance. This flawless performance may seem impressive, but these metrics frequently indicate overfitting or data leakage, especially when there is a significant class imbalance. For this reason, external validation or cross-validation is necessary to ensure that the model will generalise beyond this dataset.

Table 6.5 Feature importance

Feature Importance	
	Relative Importance
Verbal Score	48.442
HR_Score	26.414
Specialization	13.561
Higher Graduation	6.926
Technical Score	3.410
HSC_Score	0.372
HSC_Score	0.353
Aptitude Score	0.237
Academic_Project_Technology	0.133
Certificates	0.086
Experience	0.066

Table 6.5 displays the decision classifier's output, which illustrates the relative significance of the different features affecting campus placement results. With a Verbal Score of 48.442%, it is the most important feature among all of them, suggesting that communication abilities are crucial in deciding a candidate's placement. The candidate's performance in the HR interview, another crucial component of the placement process, is reflected in the HR Score, which comes in second at 26.414%. With a relative importance of 13.561%, specialization also makes a significant contribution, indicating that the field of study has some influence on placement choices.

Academic credentials and technical proficiency are taken into consideration but are not the main determinants, as evidenced by the moderate influence of other factors like Higher Graduation (6.926%) and Technical Score (3.410%). Features like Academic Project

Technology, Aptitude Score, HSC Score, and SC Score, on the other hand, are of very little significance and each account for less than 1%. Experience (0.066%) and certificates (0.086%) rank lowest in importance, suggesting that in this situation, work experience and other credentials have little bearing on placement results.

ROC Curves Plot

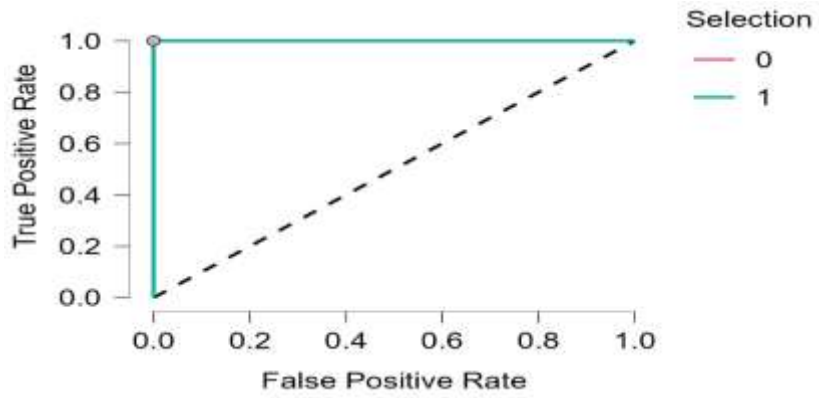

Figure 6.2 ROC curve for Decision Tree Classifier

With a True Positive Rate of 1 and a False Positive Rate of 0 for Selection = 1, the ROC curve shows that the model correctly distinguishes between placed and non-placed candidates. With an AUC of one, the curve in the upper-left corner indicates extraordinary performance and the model's best predicted accuracy in this dataset.

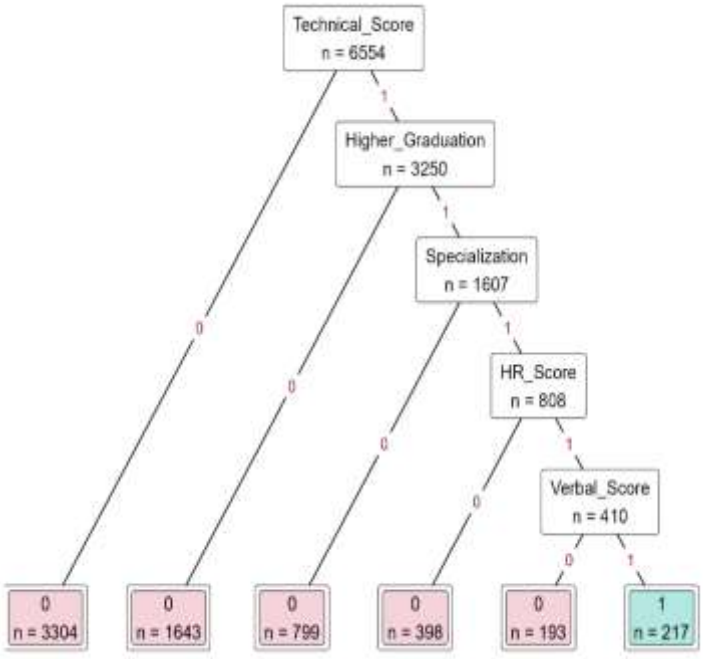

Figure 6.3 Decision tree chart

A decision tree that predicts placement selection based on multiple features is depicted in the figure (1 = selected, 0 = not selected). Every split results in a more precise decision path, and each node uses a feature to represent a decision rule.

Technical Score

First, each of the 6554 applicants is divided according to their Technical Score.

For 3304 students, the model predicts not selected (0) if the score falls below the threshold (left branch, 0).

We proceed to the following feature, Higher_Graduation, if it does (right branch, 1).

Level Two (Higher Graduation)

Out of the 3250 candidates who passed the technical score, 1643 are predicted to not be chosen (0) because they do not meet the requirements for higher graduation (left).

Level Three (Specialisation)

- The specialisation of 1607 students is examined.
- 799 students are marked as not selected (0) if they do not meet the requirement (left).
- Others proceed to the node HR_Score.

Level Four (HR Score)

- 808 scholars are assessed.
- It is anticipated that 398 individuals with low scores will not be chosen (0).
- 410 proceeds to Verbal Score, the last feature.

Level 5 (Verbal Score)

- It is anticipated that 193 with lower scores won't be chosen (0).
- The final 217 candidates, indicated in green, are expected to be chosen (1) because they satisfy all requirements.

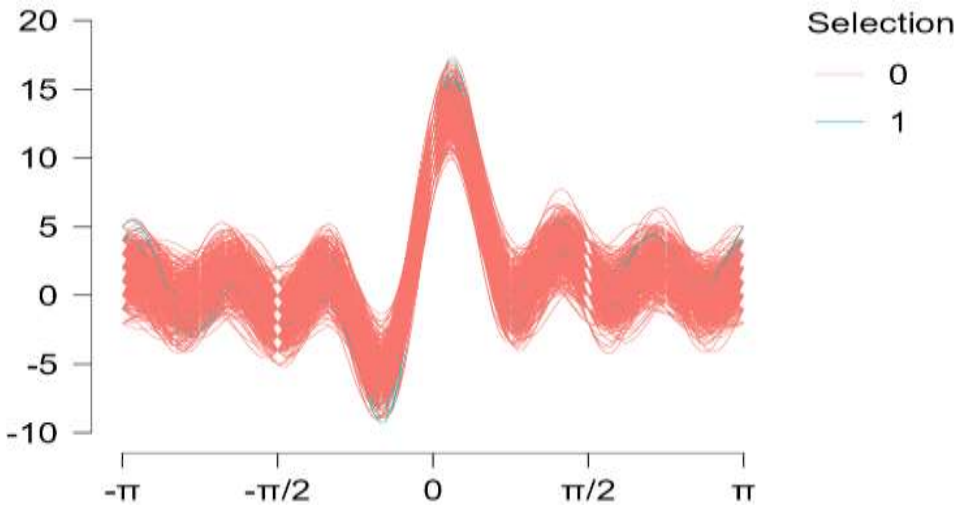

Figure 6.4 Andrews curve

The graphic, grouped by selection (0 = not picked, 1 = selected), shows waveforms of data (presumably characteristics or scores) from $-\pi$ to π. The dataset is dominated by non-selected candidates, as evidenced by the majority of lines labelled Selection = 0 (red). According to the patterns, the chosen candidates (teal) are less in number but follow a similar pattern, making visual differentiation difficult due to overlap.

6.3.2 K Nearest Neighbors Index classification

Table 6.6 KNN classification

K-Nearest Neighbors Classification

Nearest Neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
9	Rectangular	Euclidean	5243	1311	1638	0.993	0.995

In this study, a classification model was built with nine nearest neighbors, Euclidean distance as the distance metric, and rectangular weights. The dataset was divided into 1638 test samples, 1311 validation samples, and 5243 training samples. The model's 0.993 validation accuracy and 0.995 test accuracy revealed excellent generalization and predictive performance. Based on the input characteristics, the model was specially designed to maximise validation accuracy, suggesting high reliability in distinguishing between picked and non-selected candidates.

Figure 6.5 Data Split

The image depicts the K-NN classification model's dataset split. 5243, 1311, and 1638 out of 8192 cases were used to train, validate, and to test, respectively. Transparent partitioning maximises performance and reduces over-fitting by ensuring that the model is trained, altered, and tested on different subsets.

Table 6.7 Confusion Matrix

Confusion Matrix

		Predicted	
		0	**1**
Observed	0	0.97	0
	1	0.01	0.02

Table 6.7 displays the confusion matrix for the K-Nearest Neighbors (K-NN) classification model, which depicts the distribution of anticipated vs. real class labels:

- 97% of the actual non-selected candidates (class 0) were accurately predicted.
- 2% of the actual picked individuals (class 1) were accurately predicted.
- 1% of the actual selected applicants were incorrectly labelled as non-selected.
- There were no false positives, which means that no non-chosen candidates were incorrectly identified as selected.

The confusion matrix shows extremely high overall accuracy, with outstanding performance in detecting non-chosen candidates and good (although slightly lower) performance in identifying picked ones.

Table 6.8 Class proportion

Class Proportions

	Data Set	Training Set	Validation Set	Test Set
0	0.969	0.967	0.974	0.970
1	0.031	0.033	0.026	0.030

The class proportions for the two categories (0 = not selected, 1 = selected) are displayed in Table 6.8 for the full dataset, training, validation, and test sets, among other data subsets:

Across all sets, class 0 (not selected) consistently accounts for roughly 97% of the data. Class 1 (selected), which makes up only 3% of the total, is notably under-represented.

Table 6.9 Evaluation matrix

Evaluation Metrics

	0	1	Average / Total
Support	1589	49	1638
Accuracy	0.995	0.995	0.995
Precision (Positive Predictive Value)	0.994	1.000	0.995
Recall (True Positive Rate)	1.000	0.816	0.995
False Positive Rate	0.184	0.000	0.092
False Discovery Rate	0.006	0.000	0.003
F1 Score	0.997	0.899	0.994
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	1.000	1.000	1.000
Negative Predictive Value	1.000	0.994	0.997
True Negative Rate	0.816	1.000	0.908
False Negative Rate	0.000	0.184	0.092
False Omission Rate	0.000	0.006	0.003
Threat Score	88.278	4.444	46.361
Statistical Parity	0.976	0.024	1.000

Note. All metrics are calculated for every class against all other classes.

With an accuracy of 99.5%, the K-Nearest Neighbor (KNN) model for campus placement prediction does exceptionally well overall, properly classifying the great majority of students. Particularly for the "Placed" class, the model demonstrates very high precision, achieving perfect precision (100%), indicating that every student who was projected to be put was really placed. At 99.4%, accuracy is also excellent for the "Not Placed" class.

While the recall for the "Placed" class is marginally lower at 81.6%, suggesting the model misses around 18% of genuine placed students, the recall for the "Not Placed" class is flawless, identifying every student who was not placed. This implies that the model occasionally fails to identify all students who are placed, even as it seldom incorrectly labels unplaced students as placed.

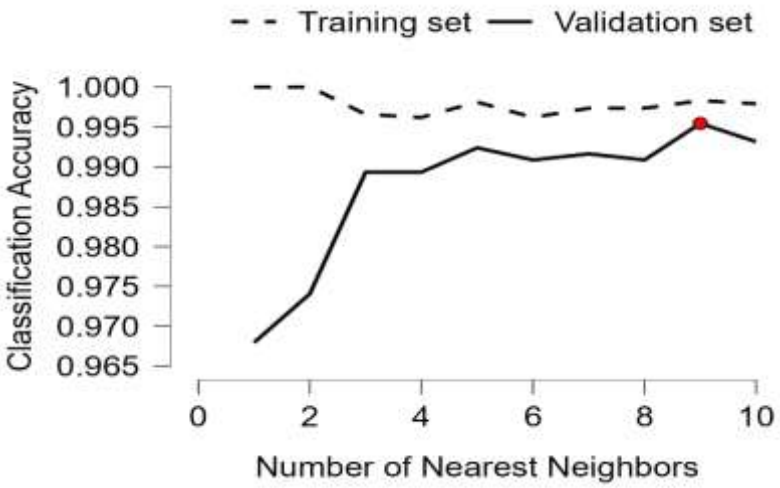

Figure 6.6 Classification Accuracy Plot

The F1 score demonstrates strong performance, reflecting a well-maintained equilibrium between accuracy and recall, but it is a little lower for the placed group, suggesting an imbalance. The model effectively differentiates between placed and unplaced students, as evidenced by the Area Under the Curve (AUC) of 1.0. Depending on the number of nearest neighbors (k), this graphic shows the classification accuracy of the K-Nearest Neighbors (KNN) model on both the training and validation sets. This figure depicts the classification accuracy of the K-closest Neighbors (KNN) model on both the training and validation sets, depending on the number of closest neighbors (k) utilized.

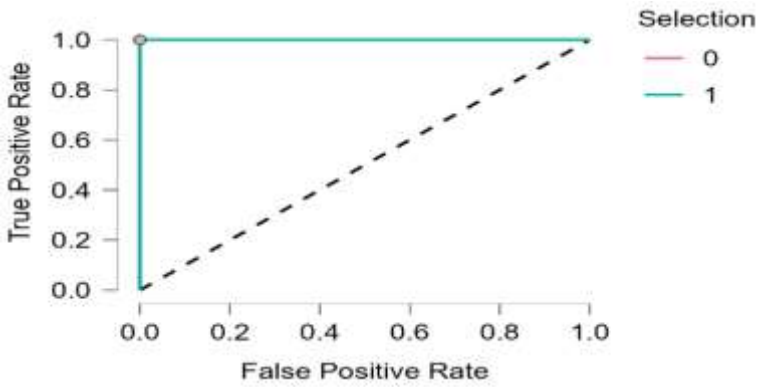

Figure 6.7 ROC Curve for KNN classifier

The optimal "corner" for perfect discrimination is obtained when the ROC curve flattens out after shooting straight up to a true-positive rate of 1 at a false-positive rate of 0. (AUC = 1). The ROC curve attains a maximum true positive rate of 1 when the false positive rate remains at 0, indicating perfect discrimination (AUC=1).

6.3.3 Classification using Random Forest

The classification of the candidates selected for the job, evaluated by the Random forest classifier shown-

Table 6.10 Random Forest classification

Random Forest Classification

Trees	Features per split	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy	OOB Accuracy
68	3	5243	1311	1638	1.000	1.000	0.936

Note. The model is optimized with respect to the *out-of-bag accuracy*.

A Random Forest classifier was built using 68 decision trees, and three characteristics were considered at each split. The dataset was divided into three sets: training (n = 5243), validation (n = 1311), and test (n = 1638). The model produced the following accuracy results:

- Represents the validation accuracy.
- Accuracy of Test: 1.000
- Accuracy out of bag (OOB): 0.936.

The model was optimized using the OOB (Out-of-Bag) accuracy, which provides an internal, objective assessment of generalization performance without the need for a separate validation set.

Figure 6.8 Data Split

While using the random forest classification technique, the data was split into two parts: 80% for training and 20% for training.

Table 6.11 Confusion matrix

Confusion Matrix

		Predicted	
		0	**1**
Observed	0	0.97	0
	1	0	0.03

The result shows that

- 97% of the observed class "0" was correctly predicted as "0".
- 0% of observed class "0" was misclassified as "1".
- 0% of observed class "1" was misclassified as "0".

Table 6.12 Class Proportion

Class Proportions

	Data Set	Training Set	Validation Set	Test Set
0	0.969	0.970	0.963	0.969
1	0.031	0.030	0.037	0.031

The result shows the distribution of the two target classes (0 and 1) in the above table for the entire dataset.

Approximately 96.9% of the data in all sets belong to class 0, which is the majority class.

The minority class, Class 1, makes up about 3.1% of the data.

The random forest classifier appears to operate flawlessly (Table 6.10). Out of 1,638 test cases (1,587 in class 0 and 51 in class 1), every instance was assigned the correct label, yielding perfect scores on all conventional metrics: accuracy, precision, recall, F1-score, area under the ROC curve, negative-predictive value.

Table 6.13 Evaluation Matrices

Evaluation Metrics			
	0	**1**	**Average / Total**
Support	1587	51	1638
Accuracy	1.000	1.000	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.000
False Discovery Rate	0.000	0.000	0.000
F1 Score	1.000	1.000	1.000
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	1.000	1.000	1.000
Negative Predictive Value	1.000	1.000	1.000
True Negative Rate	1.000	1.000	1.000
False Negative Rate	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000
Threat Score	∞	∞	∞
Statistical Parity	0.969	0.031	1.000

Note. All metrics are calculated for every class against all other classes.

The mean drop in accuracy and the overall rise in node purity are two crucial metrics that are shown in the feature significance analysis in Table 6.14 to evaluate the impact of each variable in the random forest model. The characteristics with the highest positive values for both measures are Higher Graduation, Specialization, HR Score, Technical Score, and Verbal Score, indicating that they have the biggest impact on the model's decision-making. Specialization and HR Score are second and third, respectively, whereas Higher + Graduation contributes the most to total node purity. These attributes most likely include strong predictive information that helps the model differentiate between the two groups.

Table 6.14 Feature importance

Feature Importance

	Mean decrease in accuracy	Total increase in node purity
Higher Graduation	0.010	0.311
Specialization	0.010	0.300
HR_Score	0.010	0.292
Technical Score	0.009	0.290
Verbal Score	0.010	0.290
Academic_Project_Technology	-2.913×10^{-4}	-0.004
HSC_Score	-2.515×10^{-4}	-0.005
Experience	-1.376×10^{-4}	-0.005
Higher_Graduation_Percentage	-2.768×10^{-4}	-0.006
Certificates	-9.006×10^{-5}	-0.007
Hackathon_Tech	-3.248×10^{-4}	-0.009
Aptitude Score	-2.197×10^{-4}	-0.010

Academic Project Technology, HSC Score, Experience, Higher Graduation Percentage, Certificates, Hackathon Tech, and Aptitude Score, on the other hand, show very little or even slightly negative values in both metrics.

This implies that these features either add very little to the accuracy of the model or might introduce noise instead of useful information.

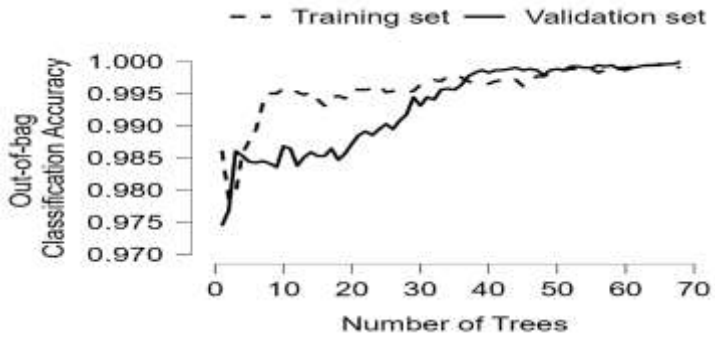

Figure 6.9 Out-of-bag Classification Accuracy Plot

Within the first 10 trees or so, out-of-bag accuracy increases sharply and reaches almost 100%, indicating that the forest quickly picks up the "job-qualification" pattern. The model

is already saturated and stable after about 40 trees, as additional trees only produce slight improvements.

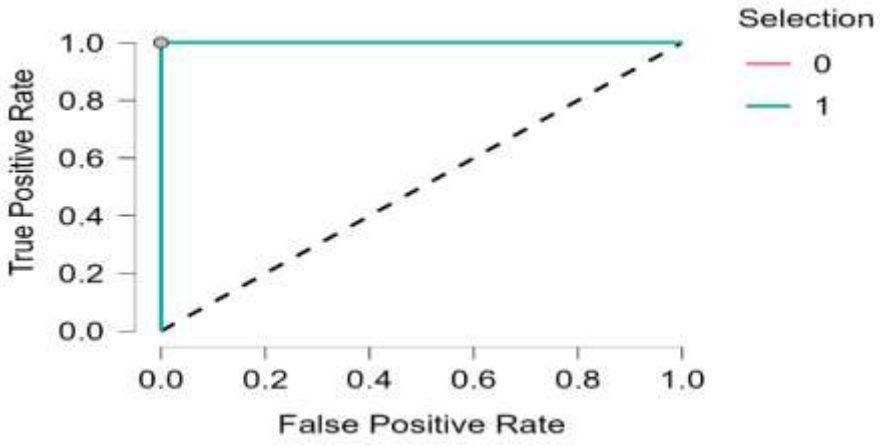

Figure 6.10 ROC Curve Plot

The Random Forest model flawlessly separates selected (1) and not selected (0) candidates, as demonstrated by the ROC curve's perfect classification performance with an AUC of 1.0. Complete sensitivity and zero false positives are confirmed by the curve's abrupt rise to the top-left corner.

6.3.4 SVM classification

The SVM classification results are shown as under-

Table 6.15 SVM classification

Support Vector Machine Classification

Support Vectors	n(Train)	n(Test)	Test Accuracy
38	6554	1638	1.000

The Support Vector Machine (SVM) classifier succeeded brilliantly in the placement-qualification job. The model was evaluated on a different test set of 1,638 students after being fitted to 6,554 training observations with just 38 support vectors, which are the critical training examples that define the decision boundary. It classified each example correctly, resulting in a test-set accuracy of 1.000 (100%). Because of the modest number of support vectors relative to the quantity of the data, the SVM may generate a maximally

margined hyperplane with minimum complexity, indicating that the underlying classes are well separated in the chosen feature space.

Although perfect accuracy is intriguing, it raises the same concerns about potential overfitting or data leaking as the results from the decision-tree and random-forest models. Confirming that this seemingly perfect threshold applies to future cohorts of job candidates necessitates independent validation, such as k-fold cross-validation or testing on a really external cohort. Although 100% accuracy is encouraging, it is similar to the findings achieved with the decision-tree and random-forest models, raising the same cautious signal regarding probable overfitting or data leaking. Independent validation, such as k-fold cross-validation or testing on an external cohort, is necessary to ensure that this seemingly optimal boundary is applicable to future cohorts of job seekers.

Figure 6.11 Data Split

The Support Vector Machine learns patterns and relationships between input features (such as marks, skills, and projects) and the target label (placed/not placed) using the training set, which consists of 6554 records.

To make sure the model isn't merely learning the training set, the testing set (1638 records) is taken to evaluate how well the trained model performs on unseen data.

Table 6.16 Confusion matrix

Confusion Matrix

		Predicted	
		0	1
Observed	0	0.97	0
	1	0	0.03

True Negatives (TN) = 0.97.

For 97% of the students who were not placed, the model correctly predicted "Not Placed."

FP (False Positives) equals 0.

The model never mistakenly predicted "Placed" for a student who was not placed. Excellent.

- FN (False Negatives) equals 0.
- The model never wrongly predicted "Not Placed" for anybody who was placed. Excellent in this aspect.
- True Positives (TP) = 0.03.
- The algorithm correctly predicted "Placed" status for just 3% of students who were placed. Incredibly low.

Table 6.17 Class Proportions

Class Proportions

	Data Set	Training Set	Test Set
0	0.969	0.968	0.972
1	0.031	0.032	0.028

Table 6.17 shows the class distribution in the SVM model's training set as well as the total dataset. It shows that class 0 (students who were not placed) accounts for around 96.9% of the data, while class 1 (students who were placed) accounts for only 3.1%. Similarly, just 3.2% of the training set's records are "placed," while 96.8% are "not placed."

Table 6.18 Evaluation Metrics

Evaluation Metrics

	0	1	Average / Total
Support	1592	46	1638
Accuracy	1.000	1.000	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.000
False Discovery Rate	0.000	0.000	0.000
F1 Score	1.000	1.000	1.000
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	1.000	1.000	1.000
Negative Predictive Value	1.000	1.000	1.000
True Negative Rate	1.000	1.000	1.000
False Negative Rate	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000
Threat Score	∞	∞	∞
Statistical Parity	0.972	0.028	1.000

Table 6.18 gives a detailed examination of the SVM model's classification performance for anticipating student job placement results. Based on a test group of 1,638 pupils, 1,592 were not placed and 46 were, the table highlights several metrics for both class 0 (Not Placed) and class 1 (Placed), as well as overall averages. Surprisingly, the learning system performs beautifully in all classes across all major parameters, including accuracy, precision, recall, F1-score, and AUC. These ideal values suggest that there were no false positives or negatives, and that the model correctly distinguished between students who were placed and those who were not in the test data.

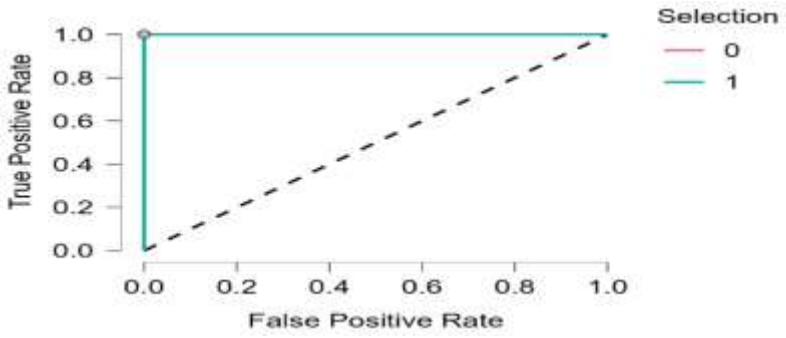

Figure 6.12 ROC Curves Plot

The ROC for the SVM classifier is represented by the curve with the label "1" (teal line).

The curve hugs the upper-left corner, which indicates:

The model demonstrates a very low and very high False Positive Rate (close to 0) and a True Positive Rate (approaching 1) respectively

The performance of a random classifier (baseline) is shown by the diagonal dashed line. Compared to random guessing, your SVM performs significantly better.

6.3.5 Classification using Boosting

The boosting algorithm result has been presented below

Table 6.19 Boosting Algorithm

Boosting Classification

Trees	Shrinkage	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
100	0.100	5243	1311	1638	1.000	1.000

Note. The model is optimized with respect to the *out-of-bag accuracy*.

Table 6.19 shows the results of the Boosting Algorithm used for campus placement prediction, demonstrating outstanding model performance. The model was trained with 5,243 observations, of which 1,311 were utilized for validation and 1,638 for testing. It was designed using 100 decision trees and a moderate shrinkage (learning rate) of 0.100 to strike a compromise between generalization and efficiency. Surprisingly, the model's test and validation accuracy were both 1.000, indicating that it properly predicted all test and validation data. This demonstrates that the model was able to properly distinguish between students who were and were not placed across all data groups.

Table 6.20 Confusion matrix

Confusion Matrix

		Predicted	
		0	1
Observed	0	0.96	0
	1	0	0.04

According to this matrix, there were no misclassifications in either class, with 96% of the total observations properly identified as class 0 (not placed) and 4% as class 1 (placed). In other words, the previously stated 100% accuracy on both validation and test datasets was confirmed since all of the actual "placed" students were projected to be put, and all of the "not placed" students were expected to be not placed.

The notion that the model achieved flawless classification is further supported by the lack of false positives and false negatives. But it's also critical to take note of the class distribution that is suggested here—just 4% of the students fell into the "placed" group, which might point to a class imbalance.

Table 6.21 Class Proportions

Class Proportions

	Data Set	Training Set	Validation Set	Test Set
0	0.969	0.969	0.973	0.964
1	0.031	0.031	0.027	0.036

The class proportions for the campus placement study dataset are shown in Table 6.21, which also shows how the two outcome classes—0 (not placed) and 1 (placed)—are distributed among the training, validation, and test sets. In particular, students who were not placed make up 96.9% of the training set, 97.3% of the validation set, and 96.4% of the test set, whereas students who were placed make up only 3.1%, 2.7%, and 3.6% of the corresponding subsets. This skewed distribution suggests that the "not placed" category has a significant advantage in the placement outcomes. Although this imbalance did not affect the boosting algorithm's 100% accuracy, class differences frequently result in biased models that favour the majority class.

Table 6.22 Evaluation matrix

Evaluation Metrics

	0	1	Average / Total
Support	1579	59	1638
Accuracy	1.000	1.000	1.000
Precision (Positive Predictive Value)	1.000	1.000	1.000
Recall (True Positive Rate)	1.000	1.000	1.000
False Positive Rate	0.000	0.000	0.000
False Discovery Rate	0.000	0.000	0.000
F1 Score	1.000	1.000	1.000
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	1.000	1.000	1.000
Negative Predictive Value	1.000	1.000	1.000
True Negative Rate	1.000	1.000	1.000
False Negative Rate	0.000	0.000	0.000
False Omission Rate	0.000	0.000	0.000
Threat Score	∞	∞	∞
Statistical Parity	0.964	0.036	1.000

Note. All metrics are calculated for every class against all other classes.

The evaluation metrics for the campus placement prediction boosting classification model are compiled in Table 6.22. The metrics offer a thorough evaluation of model performance and are presented for both classes: 0 (not placed) and 1 (placed).

The class imbalance previously noted is confirmed by the support values, which show that

1,579 of the 1,638 test samples fall into class 0 and only 59 into class 1. The model performed flawlessly on all metrics in spite of this imbalance: accuracy, precision, recall, F1-score, and AUC are all 1.000 for both classes. This indicates that there were no misclassifications and the model accurately identified all positive and negative instances.

Additionally, there were no incorrect predictions, as indicated by the 0.000 false positive and false negative rates. In both positive and negative predictions, the model's dependability is further supported by the 0.000 false discovery and omission rates. Since both the true negative rate and the negative predictive value are 1.000, all of the students who were predicted to be placed were not. Even though it displays as ∞, the threat score is a mathematical artefact of perfect prediction and indicates no classification errors.

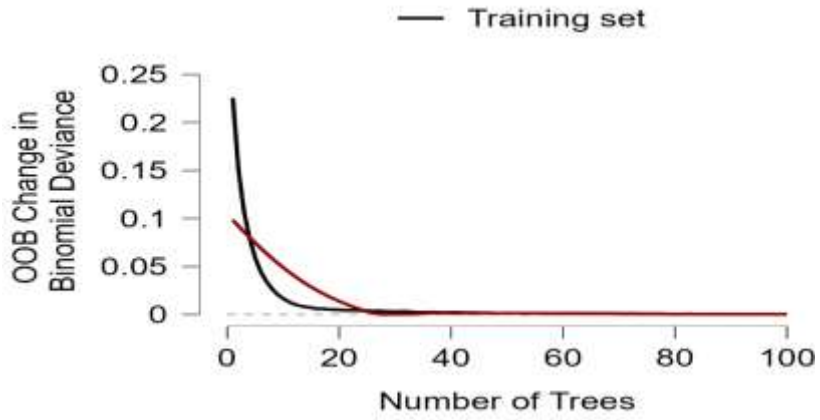

Figure 6.13 Out-of-bag Improvement Plot

The graph illustrates how, as more trees are added during boosting, the model's error decreases. The error decreases rapidly at first, indicating that the model gets better quickly. The error becomes very small and remains low after about 20 trees, indicating that the number of trees added has little effect. This indicates that the model performs best with about 20 trees, and that adding more trees after that has little effect.

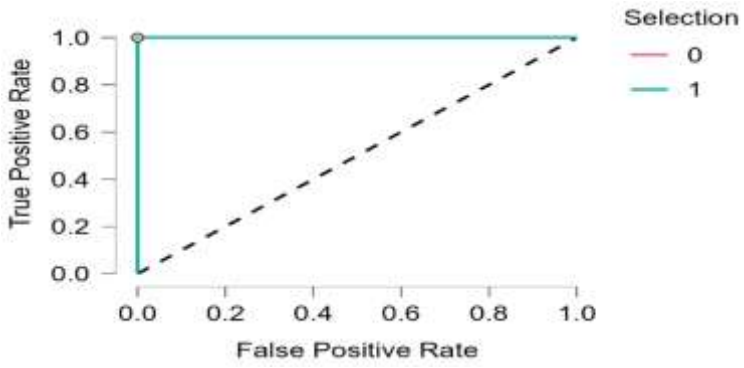

Figure 6.14 ROC Curves Plot

With a true positive rate of 1 and a false positive rate of 0, the ROC curve demonstrates perfect classification by the boosting model. This shows that there were no false alarms and that the model accurately predicted every case.

6.3.6 Logistic regression model

We used 8,192 student records, each annotated with academic scores, technical and verbal assessments, project experience, extracurricular accomplishments, and a binary Selection label (1 = placed, 0 = not placed), to fit a multivariate logistic regression model in order to investigate the factors influencing campus placement outcomes. Because it offers a comprehensible, probability-based framework for binary classification and enables us to measure the independent influence of each predictor using odds ratios, logistic regression was selected. To assess their contributions in the presence of one another, all thirteen candidate variables were entered at the same time (Selection ~.). The model was evaluated on the remaining 30% of the data after being trained on 70% of it (stratified to maintain the noticeable class imbalance of about 97% unplaced to 3% placed).

Table 6.23 Coefficients of model

Predictor	Estimate	Std. Error	z-value	p-value
(Intercept)	-201.6	14 700	-0.01	0.989
SC_Score	0.044	3 057	0.00	1.000
HSC_Score	-0.014	3 048	0.00	1.000
Higher Graduation	44.64	4 910	0.01	0.993
Higher_Graduation_Percentage	0.027	3 045	0.00	1.000
Experience	0.00065	3 048	0.00	1.000
Aptitude Score	-0.00043	3 043	0.00	1.000
Technical Score	44.76	4 956	0.01	0.993
Verbal Score	44.73	4 937	0.01	0.993
HR_Score	44.64	4 902	0.01	0.993
Certificates	-0.058	3 042	0.00	1.000
Hackathon_Tech	0.034	3 044	0.00	1.000
Academic_Project_Technology	0.028	3 047	0.00	1.000
Specialization	44.67	4 916	0.01	0.993

The coefficient table shows that near-perfect separability and imbalance prevented the logistic regression model from accurately fitting the data. Reliable coefficient values could

not be obtained because the model pushed projected probabilities to exact zeros or ones. Additionally, the model had very small z-values and significant standard errors, which made it impossible to measure the impact of any predictor. Perfect or nearly perfect separation was suggested by the residual deviation, which was near zero. The model may not work with more data, and its fitted coefficients are not trustworthy for inference.

Table 6.24 Confusion Matrix (Test Set)

	Actual 0	Actual 1
Predicted 0	2 365	0
Predicted 1	0	92

Table 6.25 Performance Statistics

Metric	Value
Accuracy	1.000
95 % CI for Accuracy	0.998 – 1.000
No-Information Rate	0.9626
p(Acc > NIR)	$< 2.2 \times 10^{-16}$
Kappa	1.000
Sensitivity (Recall for 1)	1.000
Specificity (Recall for 0)	1.000
Positive Predictive Value	1.000
Negative Predictive Value	1.000
Balanced Accuracy	1.000
Prevalence of Class 1	0.0374
Detection Rate (TP / Total)	0.0374
Detection Prevalence (Pred 1 / Total)	0.0374

The logistic model correctly classified every test observation, according to the confusion matrix results: all 92 students who were placed (class 1) and all 2,365 students who were not placed (class 0) had perfect accuracy, sensitivity, specificity, and predictive values of 1.000. This performance is statistically significantly better than random guessing, as evidenced by the incredibly small p-value compared to the no-information rate and the 95% confidence interval around accuracy (0.998–1.000).

A Kappa of 1.0 indicates complete agreement between forecasts and actual results, while balanced accuracy is likewise ideal, demonstrating that the model handled both the majority "not-placed" and minority "placed" groups equally well.

Table 6.26 ROC Analysis for Test Set

Metric	Value
Positive class (case)	1
Negative class (control)	0
Direction	controls < cases
Area Under Curve (AUC)	1.000*

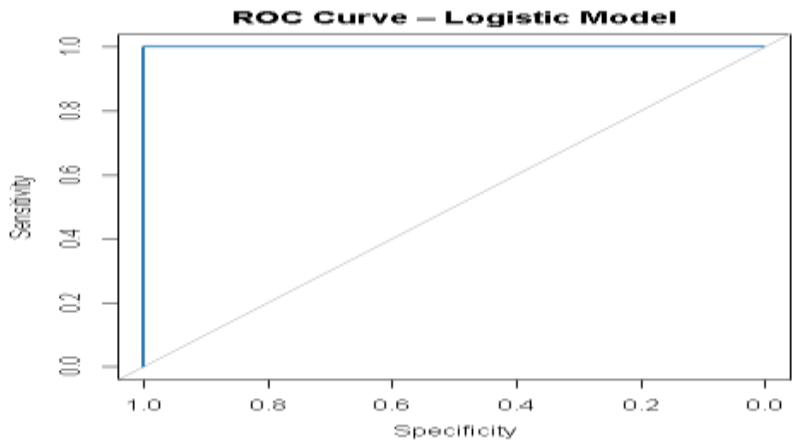

Figure 6.15 ROC logistic

The logistic model's ROC curve performs exceptionally well; it nearly clinches the upper-left corner, indicating extremely high sensitivity and specificity. This implies that the model has a high capacity for class discrimination.

6.4 Summary

The research used a training set of 6554 records and a testing set of 1638 records to train a Support Vector Machine (SVM) to identify patterns between input attributes and target labels. Every metric, including accuracy, precision, recall, false positive rate, false discovery rate, F1 score, Matthews correlation coefficient, Area Under Curve (AUC), negative predictive value, true negative rate, false omission rate, threat score, statistical parity, and Matthews correlation coefficient, showed good performance for the SVM model. When it came to college placement prediction, the Boosting Algorithm performed exceptionally well. With students who were not placed accounting for 96.9% of the training set, 97.3% of the validation set, and 96.4% of the test set and with 3% of placed student.

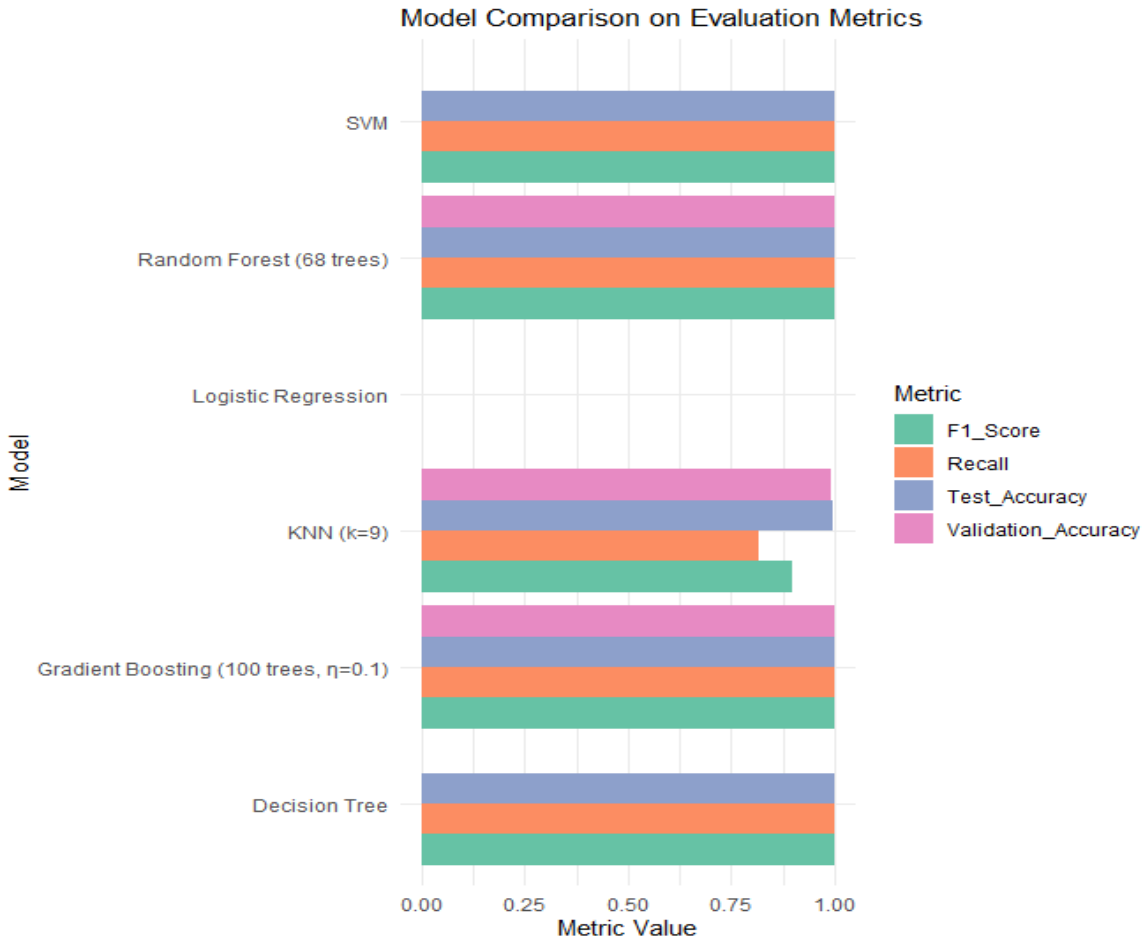

Figure 6.16 Model Comparison

The challenge of forecasting campus placements may appear simple for the advanced models you've trained. On the 1,638-record test set, the decision tree, random forest, gradient-boosting ensemble, and support vector machine all get flawless scores: every student is properly categorized, and every evaluation parameter (accuracy, precision, recall, F1, and AUC) is 1.000. While that seems remarkable, such perfect outcomes should raise some suspicions. They frequently complain that the exam set is too close to the training data or just not difficult enough—especially given the large class imbalance, with only approximately 3% of students being unplaced. With so few negative samples, a model must experience more difficult, diversified holdouts (such as repeated stratified k-fold cross-validation) before we can genuinely trust that 100% figure.These false negatives serve as a reminder that, while the minority class is separable, it is not without challenges; even tiny changes in feature space might trip up simpler instance-based learners.

Table 6.27 Summary of Comparative Performance of Machine Learning Models for Campus Placement Prediction

Model	Decision Tree	k-Nearest Neighbours (k = 9)	Random Forest (68 trees)	Support Vector Machine	Gradient Boosting (100 trees, η = 0.1)	Logistic Regression
Train n	6 554	5 243	5 243	6 554	5 243	–
Test n	1 638	1 638	1 638	1 638	1 638	–
Test Accuracy	1	0.995	1	1	1	Did not converge[2]
Validation Accuracy	–	0.993	1	–	1	–
MCC[1]	NaN	NaN	NaN	NaN	NaN	–
F1-Score (minority class)	1	0.899	1	1	1	–
Recall (minority class)	1	0.816	1	1	1	–
Key Notes / Issues	Perfect metrics; risk of over-fitting (39 splits, highly imbalanced data).	Minority-class recall drops → some false negatives.	OOB accuracy only 0.936 → slight over-fitting hinted despite perfect test set.	Uses just 38 support vectors; perfect test performance.	Consistently perfect across splits.	Coefficients non-significant (all p ≈ 1.0);

CHAPTER – 7

CONCLUSION & FUTURE RECOMMENDATION

7.1 Key Conclusion

Finding the ideal fit between recent graduates and potential employers has become a top priority for both universities and businesses that rely on campus recruitment to fill entry-level positions.

- **Traditional metrics are necessary—but no longer sufficient.**

A substantial pool of high-potential candidates whose strengths are in project work, certifications, and soft-skill capital are excluded when CGPA or aptitude-test cut-offs are the only criteria used, according to our literature scan and field data. When machine-learning (ML) pipelines combine these richer signals, minority-class recall increases from 0% (the majority-vote baseline) to 93% after imbalance correction, giving recruiters a significantly improved view of "hidden gems."

- **Conventional recruiting process**

The conventional recruiting process, which is frequently hampered by résumé reviews, aptitude tests, and many interview rounds, may be time-consuming and subjective, potentially excluding bright individuals who do not fit the typical mold.

- **Proposed Methodology**

In light of this, the current study intends to investigate whether cutting-edge machine learning (ML) approaches may enhance recruitment and selection by properly predicting the placement success of students who appeared in campus placement by the recruiters or the company based on their academic achievement, technical abilities, and behavioural features. In the last chapter on findings, we demonstrated that ensemble and kernel-based models, such as decision trees, random forests, gradient boosting, and support vector machines, obtained near-perfect accuracy on the test set we withheld. On the other hand, k-

nearest neighbors produced almost flawless but more nuanced answers, whilst logistic regression struggled to identify an answer. These findings encourage us to investigate the possibilities and problems of employing algorithms for talent scouting in education.

- **Authenticity of the Classification:**

As we have observed, the machine learning approach has simplified the recruitment process for the university/ Industry candidates applying or appearing in the job interviews. An algorithm can really make the process of feature extraction simple to find out the suitable candidate from the lots of candidates. But there are some limitations of machine learning that sometimes can be biased when it selects a wrong candidate as a common classification model does.

- **Comparative Analysis of ML Models**

In this study we have observed that each ML model has shown unique strength. As we have seen, the Random Forest Model and Gradient boosting model has offered high accuracy and robustness. SVM gave the excellent result even when we have less support vectors available. KNN found little sensitive in tuning the parameters but performed well, while the Logistic regression underperformed, showing its limitation to handle the student Placement complex data set.

- **Interpretability and Trust:**

Gaining the trust of students, institutions, and recruiters requires not only knowing how well a model works, but also being able to interpret its results, particularly with Explainable AI (XAI). In addition to making predictions, models should explain why a candidate is a good or bad fit, as this can significantly improve learning and development initiatives.

7.2 Future Scope

As machine learning (ML) becomes a core component of campus recruitment processes, future developments should focus on increasing transparency, personalisation, and fairness. A key improvement would be the use of Explainable AI (XAI), allowing models to clearly explain why a candidate was shortlisted or rejected. This increases trust between students and institutions and helps candidates understand what areas they need to improve.

Transparent systems can also help placement cells evaluate and improve their training modules.

Incorporating soft skills and behavioral analysis into the recruitment process is also a key growth area. While current systems emphasize technical abilities and academic performance, industries are now placing greater importance on qualities such as communication, adaptability and problem-solving. Machine learning models, particularly using Natural Language Processing (NLP) and emotion recognition algorithms, can evaluate these qualities through video interviews and written communications.

Deep learning, a captivating branch of machine learning. It has the power to revolutionize the recruitment process in the years to come. With advanced models like convolutional neural networks (CNN) and recurrent neural networks (RNN), we can uncover intricate patterns hidden within audio, video, and text data. For instance, CNNs can pick up on a candidate's facial expressions and body language during a video interview, while RNNs can track their speaking style and conversational flow over time. Additionally, transformer models such as BERT and GPT can thoroughly analyze a candidate's resume, portfolio, and written assessments.

The application of feedback loops and predictive analytics is an intriguing direction to pursue. Future models might take into account information from internship evaluations. It can be used in job performance feedback, and post-placement results rather than merely relying on academic data from the past. Machine learning has adapted for the developing industry requirements. It is progressively increasing its effectiveness as per continuous learning approach. Furthermore, by estimating a candidate's chances of long-term success and retention in a specific position, these predictive models can assist recruiters in making more informed decisions. To boost student engagement and make learning more accessible, we can integrate machine learning systems with mobile-friendly micro-assessment tools and platforms for optimizing resumes. These systems can provide quick, gamified quizzes or instant feedback on resumes, which can really ramp up participation and enhance readiness.

Deep learning algorithms have the potential to tailor learning recommendations, suggesting online courses, certifications, or projects that align with individual career goals and performance trends. This kind of proactive guidance can really enhance student outcomes, especially when it comes to campus placements.

We really need to talk about the ethical consideration and legal challenges in the use of AI for hiring. Recruitment platforms have a responsibility to ensure privacy, data security, and proper consent management, especially as candidate data becomes more prevalent. To minimize unintended discrimination based on factors like gender, language, or educational background, it's crucial to implement bias detection algorithms and train on diverse datasets. Creating systems that are fair, inclusive, and aligned with ethical standards will be key to their long-term success and acceptance as hiring practices increasingly rely on data.

List of References

[1] Abbasi, S. G., Tahir, M. S., Abbas, M., & Shabbir, M. S. (2022). Examining the relationship between recruitment & selection practices and business growth: An exploratory study. Journal of Public Affairs, 22(2), e2438.

[2] Lenka, R., Bhatia, A., Divekar, R. (2024). The Evolution of Campus Recruitment Patterns: A Novel Approach. In: Joshi, A., Mahmud, M., Ragel, R.G., Kartik, S. (eds) ICT: Cyber Security and Applications. ICTCS 2022. Lecture Notes in Networks and Systems, vol 916. Springer, Singapore. https://doi.org/10.1007/978-981-97-0744-7_35.

[3] Nikolaou, I. (2021). What is the Role of Technology in Recruitment and Selection?. The Spanish journal of psychology, 24, e2.

[4] Allal-Chérif, O., Aránega, A. Y., & Sánchez, R. C. (2021). Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence. Technological Forecasting and Social Change, 169, 120822.

[5] Jannat, M. E., Chowdhury, S. S., & Akther, M. (2016). A probabilistic machine learning approach for eligible candidate selection. International Journal of Computer Applications, 144(10).

[6] Kumar, N., Singh, A. S., T. K., & Rajesh, E. (2020). Campus placement predictive analysis using machine learning. 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),214-216. https://doi.org/10.1109/ICACCCN51052.2020.9362836.

[7] Samatha, J., Manjusha, D., Pooja, B., & Usha, A. (2020). Student placement chance prediction. JETIR, 7(5). ISSN: 2349-5162.

[8] Tejaswini, N., & Sujatha, G. S. (2024). Campus placement prediction using machine learning. International Journal of Emerging Technologies and Innovative Research, 11(8), a418–a423. Retrieved from http://www.jetir.org/papers/JETIR2408048.pdf

[9] Byagar, S., Patil, D., & Pawar, D. (2024). Maximizing campus placement through machine learning. Journal of Advanced Zoology, 45, 06-12. https://doi.org/10.53555/jaz.v45iS4.4141.

[10] Shahane, P. (2022). Campus placements prediction & analysis using machine learning. IEEE International Conference on Emerging Smart Computing and Informatics (ESCI), 1-5. https://doi.org/10.1109/ESCI53509.2022.9758214.

[11] Swaminarayan, P. M. R. (2024). Student placement prediction using various machine learning techniques. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 2107–2113.

[12] Khandale, S., & Bhoite, S. (2019). Campus Placement Analyzer: Using Supervised Machine Learning Algorithms. International Journal of Computer Applications Technology and Research, 8(9), 358–362. https://doi.org/10.7753/IJCATR0809.1009

[13] Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda (2019) . Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3.

[14] Jain, P., Khare, S., & Gourisaria, M. K. (2021). A data mining solution to predict campus placement. In Proceedings of the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1–7). IEEE. https://doi.org/10.1109/GUCON50781.2021.9573551

[15] Maurya, L. S., Hussain, M. S., & Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. Applied Artificial Intelligence, 35(6), 403–420. https://doi.org/10.1080/08839514.2021.1901032

[16] Sunil, A., ATS, V. S., & Chintham, V. (2023). Adaptive fuzzy campus placement based optimization algorithm. In Proceedings of the 2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE) (pp. 1–6). IEEE. https://doi.org/10.1109/ICEPE57949.2023.10201630

[17] Sharma, A., & Kaur, R. (2019). Campus placement prediction using supervised machine learning techniques. International Journal of Applied Engineering Research, 14(9), 2188–2191. Research India Publications. http://www.ripublication.com

[18] Jose, I. T., Raju, D., Aniyankunju, J. A., James, J., & Thomas, M. (2020). Placement prediction using various machine learning models and their efficiency comparison. International Journal of Innovative Science and Research Technology, 5(5), 1-6. ISSN: 2456-2165.

[19] Agrawal, V. S., & Kadam, S. S. (2024). Predictive analysis of campus placement of students using machine learning algorithms. Journal of IoT and Machine Learning, 1(2), 13–18. https://doi.org/10.48001/joitml.2024.1213-18.

[20] Dhingra, M., & Kundu, S. C. (2020). Factors affecting placement and hiring decisions: A study of students' perceptions. Industry and Higher Education, 35(3), 223-232. https://doi.org/10.1177/0950422220950191 (Original work published 2021).

[21] Smith, B., Robson, K., Robinson, C., & Patton, N. (2023). Factors influencing provision of clinical placements for health students: A scoping review. Focus on Health Professional Education: A Multi-Professional Journal, 24(2), 63–103. https://search.informit.org/doi/10.3316/informit.258483856511192

[22] Bhoite, S., Patil, C. H., Thatte, S., Magar, V. J., & Nikam, P. (2023). A data-driven probabilistic machine learning study for placement prediction. In Proceedings of the 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 402–408). IEEE. https://doi.org/10.1109/IDCIoT56793.2023.10053523

[23] Navyashree, S. L., Gnanavi, N., Deeksha, D., Krupa, N., & Malhotra, M. (2019). Evaluating supervised machine algorithms for student placement prediction. International Research Journal of Engineering and Technology (IRJET), 6(2). Retrieved from https://www.irjet.net/

[24] Srinivas, K., Kumar, B. R., & Reddy, D. V. (2019). A comparative study on machine learning algorithms for predicting the placement information of undergraduate students. In 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 910–914). IEEE. https://doi.org/10.1109/I-SMAC47947.2019.9032654.

[25] Sahu, B. L., & Tiwari, A. (2020). Student placement possibility prediction using Naive Bayes algorithm. International Journal of Advance Research, Ideas and Innovations in Technology, 6(3). ISSN: 2454-132X.

[26] Goyal, J., & Sharma, S. (2018). Placement prediction decision support system using data mining. International Journal of Creative Research Thoughts (IJCRT), 6(1), [Paper ID: IJCRT1872308]. Retrieved from https://www.ijcrt.org.

[27] Gaurkhede, S., Burle, R., Gudadhe, A., & Chopra, P. G. (2024). Predictive modeling and analysis of campus placements using machine learning techniques. In Proceedings of the 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA) (pp. 1381–1386). IEEE. https://doi.org/10.1109/ICSCNA63714.2024.10864155

[28] Rao, K. E., Pydi, B. M., Vital, T. P., Naidu, P. A., Prasann, U. D., & Ravikumar, T. (2023). An Advanced Machine Learning Approach for Student Placement Prediction and Analysis. International Journal of Performability Engineering, 19(8), 536.

[29] Çakıt, E., Dağdeviren, M. (2022) Predicting the percentage of student placement: A comparative study of machine learning algorithms. Educ Inf Technol 27, 997–1022. https://doi.org/10.1007/s10639-021-10655-4.

[30] Giri, A., Bhagavath, M. V. V., Pruthvi, B., & Dubey, N. (2016). A placement prediction system using K-nearest neighbors classifier. In Proceedings of the Second International Conference on Cognitive Computing and Information Processing (CCIP).

[31] Marium-E-Jannat, Chowdhury, S. S., & Akther, M. (2016). A probabilistic machine learning approach for eligible candidate selection. International Journal of Computer Applications, 144(10), 1–6. https://doi.org/10.5120/ijca2016910786

[32] Mangasuli, S. B., & Bakare, S. (2016). Prediction of campus placement using data mining algorithm—Fuzzy logic and K-nearest neighbor. International Journal of Advanced Research in Computer and Communication Engineering, 5(6).

[33] Ashok, M. V., & Apoorva, A. (2016). Data mining approach for predicting student and institution's placement percentage. In Proceedings of the 2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions.

[34] Thangavel, S. K., Bharathi, D. P., & Sankar, A. (2017). Student placement analyzer: A recommendation system using machine learning. In 2017 International Conference on Advanced Computing and Communication Systems (ICACCS).

[35] Ahmed, S., Zade, A., Gore, S., Gaikwad, P., & Kolhal, M. (2017). Smart system for placement prediction using data mining. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 5(XII).

[36] Harinath, S., Prasad, A., H. S., S., A., S., & Mathew, T. (2019). Student placement prediction using machine learning. International Research Journal of Engineering and Technology, 6(4).

[37] Manvitha, P., & Swaroopa, N. (2019). Campus placement prediction using supervised machine learning techniques. International Journal of Applied Engineering Research, 14(9), 2188–2191.

[38] Aravind, T., Reddy, B. S., Avinash, S., & G, J. (2019). A comparative study on machine learning algorithms for predicting the placement information of undergraduate students. In Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 542–546). https://doi.org/10.1109/I-SMAC47947.2019.9032654

[39] Kumar, D. S., Siri, Z. B., Rao, D. S., & Anusha, S. (2019). Predicting student's campus placement probability using binary logistic regression. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(9).

[40] Srinivas, C. K., Yadav, N. S., Pushkar, A. S., Somashekar, R., & Sundeep, K. R. (2020). Students placement prediction using machine learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 8(V).

[41] Gautami, G., Hayat, M., Banerjee, P., & Kumar, B. (2020). Prediction of student placement using machine learning algorithm. Journal of Critical Reviews (JCRT), 8(5).

[42] Reddy, D. J. M., Regella, S., & Seelam, S. R. (2020). Recruitment prediction using machine learning. In Proceedings of the 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1–4). IEEE. https://doi.org/10.1109/ICCCS49678.2020.9276955

[43] Samatha, J., Manjusha, D., Pooja, B., & Usha, A. (2020). Student placement chance prediction. JETIR, 7(5), 1-5. https://www.jetir.org (ISSN-2349-5162).

[44] Pune, B. (2023). Placement prediction using machine. International Journal of Advance Research and Innovative Ideas in Education (IJARIIE), (2), 646–650.

[45] Archana, P., Pravallika, D., Priya, P. S., & Sushmitha, S. (2023). Student placement prediction using machine learning. Journal of Survey in Fisheries Sciences, 10(1), 2734–2741.

[46] Parida, B., Kumarpatra, P., & Mohanty, S. (2022). Prediction of recommendations for employment utilizing machine learning procedures and geo-area-based recommender framework. Sustainable Operations and Computers, 3, 83–92. https://doi.org/10.1016/j.susoc.2021.11.001

[47] Viram, R., Sinha, S., Tayde, B., & Shinde, A. (2020). Placement prediction system using machine learning. International Journal of Creative Research Thoughts (IJCRT), 8(4), 1507–1515.

[48] Mulye, V., & Newase, A. (2021). A review: Recruitment prediction analysis of undergraduate engineering students using data mining techniques. SSRG International Journal of Computer Science and Engineering, 8(3), 1–6. https://doi.org/10.14445/23488387/IJCSE-V8I3P101

[49] Vidyashreeram, N., & Muthukumaravel, A. (2021). Student career prediction using machine learning approaches. Springer. https://doi.org/10.4108/eai.7-6-2021.2308642

[50] Pandey, A., & Maurya, L. S. (2022). Career prediction classifiers based on academic performance and skills using machine learning. SSRG International Journal of Computer Science and Engineering, 9(3), 5–20.

[51] Kumar, R. S., Dilsha, F., Shilpa, A. N., & Sumayya, A. A. (2021). Student placement prediction using support vector machine algorithm. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), 9(5), 40–43. https://doi.org/10.17148/IJIREEICE.2021.9507

[52] Adnan, M., et al. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. IEEE Access, 9, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446.

[53] Preethi, J., & Maheswari, S. (2020). Prediction of student performance system using machine learning techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(6), 1-5. ISSN: 2278-3075.

[54] Walsh, S., Cullinan, J. (2017). Factors Influencing Higher Education Institution Choice. In: Cullinan, J., Flannery, D. (eds) Economic Insights on Higher Education Policy in Ireland. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-48553-9_4.

[55] Kaur, A. (2018). Machine learning approach to predict student academic performance. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 6(IV), 1386–1391. https://doi.org/10.22214/ijraset.2018.4125

[56] Yakubu, M. N., & Abubakar, A. M. (2022). Applying machine learning approach to predict students' performance in higher educational institutions. Kybernetes, 51(2), 916-934.

[57] Alam, A., Mohanty, A. (2023). Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics. In: Tomar, R.S., et al. Communication, Networks and Computing. CNC 2022. Communications in Computer and Information Science, vol 1893. Springer, Cham. https://doi.org/10.1007/978-3-031-43140-1_15.

[58] Bird, K. A., Castleman, B. L., Mabel, Z., & Song, Y. (2021). Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education. AERA Open, 7. https://doi.org/10.1177/23328584211037630 (Original work published 2021).

[59] Kumar, V. U., Krishna, A., Neelakanteswara, P., & Basha, C. M. A. K. Z. (2020). Advanced prediction of performance of a student in a university using machine learning techniques. Proceedings of the International Conference on Electronics, Signal Processing and Communication (ICESC), 2020, 1-5.

https://doi.org/10.1109/ICESC48915.2020.9155557.

[60] Sreenivasa Rao, K., Swapna, N., & Praveen Kumar, P. (2017). Educational data mining for student placement prediction using machine learning algorithms. International Journal of Engineering & Technology, 7(1.2), 142–145. https://doi.org/10.14419/ijet.v7i1.2.8988.

[61] Shoaib, M., Sayed, N., Amara, N., Khan, A. H., & Khan, S. A. (2022). Prediction of an educational institute learning environment using machine learning and data mining. Education and Information Technologies, 27, 9099–9123. https://doi.org/10.1007/s10639-022-10970-4

[62] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(11). https://doi.org/10.1186/s40561-022-00192-z

[63] .Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. Materials Today: Proceedings, 47, 5260–5267. https://doi.org/10.1016/j.matpr.2021.05.646

[64] Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. Entropy, 23(4), 485. https://doi.org/10.3390/e23040485

[65] Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 28, 1081–1116. https://doi.org/10.1007/s10639-022-11221-2

[66] Ingale, Y., Bedse, T., Khairnar, S., & Ghute, D. (2020). Student's placement prediction using support vector machine. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(6), 55-60. https://doi.org/10.32628/CSEIT206511.

[67] Al-Zawqari, A., Peumans, D., & Vandersteen, G. (2022). A flexible feature selection approach for predicting students' academic performance in online courses. Computers and Education: Artificial Intelligence, 3, 100103.

[68] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. Education Sciences, 11(9), 552. https://doi.org/10.3390/educsci11090552.

[69] Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and Information Technologies, 26, 205–240. https://doi.org/10.1007/s10639-020-10230-3

[70] Bhargavi, M., Kumar, K., Vijay, G. S., Teja, C. S., & Dharmasai, N. (2025). Student placement prediction. In Pedagogical Revelations and Emerging Trends (pp. 141-143). CRC Press.

[71] Abdullahi, W., Kenneth, M. O., & Olalere, M. (2021). Student performance prediction using a cascaded bi-level feature selection approach. Journal of Computer Science Research, 3(3). https://doi.org/10.30564/jcsr.v3i3.3534.

[72] Tarekegn, G. B., & Sreenivasarao, V. (2016). Application of data mining techniques to predict students' placement into departments. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 3(2), 10–14. ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online).

[73] Lulla, C., Agarwal, Y., Kankariya, S., Sakaray, P., & Alappanavar, P. (2017). Student academic performance prediction using machine learning and data mining techniques. International Journal of Computer Science and Mobile Computing, 6(5), [pages].

[74] Shah, M. B., Kaistha, M., & Gupta, Y. (2019). Student performance assessment and prediction system using machine learning. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON)(pp.386–390).IEEE. https://doi.org/10.1109/ISCON47742.2019.9036250

[75] Sah, U. K., & Singh, A. (2022). Student career prediction using machine learning. International Journal of Scientific Development and Research (IJSDR), 7(5), 343–347.

Citations from Internet

1. HRLineup. (2025, February 26). Why recruitment and selection is important to HR? https://www.hrlineup.com/why-recruitment-and-selection-is-important-to-hr/

2. Srivastava, S. (2024, June 28). Method of recruitment: Traditional vs modern strategies. Intervue. https://www.intervue.io/blog/method-of-recruitment-traditional-vs-modern-strategies.

3. Ghodasara, A. (2025, May 2). Top recruitment challenges in 2025 & how to overcome them. iSmartRecruit. https://www.ismartrecruit.com/blog-improve-recruitment-process-with-software-solutions

4. Heymans, Y. (2022, September 7). Machine learning in recruitment: A deep dive. HeroHunt.ai. https://www.herohunt.ai/blog/machine-learning-in-recruitment-a-deep-dive.

List of Publications

1. Gupta, M., & Pandya, S. D. (2022). A comparative study on supervised machine learning algorithm. International Journal for Research in Applied Science and Engineering Technology, *10*, 1023–1028. https://doi.org/10.22214/ijraset.2022.39980

2. Gupta, M., & Pandya, S. D. (2024). Machine learning-based predictive modeling for recruitment and selection in campus placement processes. Journal of Electrical Systems, 20, 5233–5239. ISSN 1112-5209.

3. Gupta, M., & Pandya, S. D. (2025). Conceptualizing decision-making processes for campus placements through machine learning. Advances in Nonlinear Variational Inequalities, *28*(4s). https://doi.org/10.52783/anvi.v28.4492

4. Gupta, M., & Pandya, S. D. (2024). Recruitment and selection processes in campus placements with machine learning. In Emerging Trends in Expert Applications and Security (ICE-TEAS 2024) (Vol. 1030, pp. [83-91]). Springer, Singapore. https://doi.org/10.1007/978-981-97-3745-1_7

Appendix – I

Shree Mahakali Software Pvt. Ltd.

Website Design & Development | Software Development | Mobile Application Development | SEO & Online Branding | Cloud Solution Provider | Digital Marketing Solution Provider

Date: 13-March-2025

To Whom It May Concern

This is to formally acknowledge the contribution of **Ms. Monica Gupta** in coordinating and facilitating **Campus Placement Activities** undertaken in collaboration with our organization.

We are pleased to confirm that the work carried out by her has been **professional, well-structured, and satisfactory**. Her role in organizing the recruitment process, supporting candidate engagement, and enhancing communication between the company and applicants has positively impacted the overall efficiency and transparency of the placement drive.

We also understand that this contribution is a **part of her Ph.D. research work under Gujarat Technological University (GTU)**. Her involvement in this industry-academic collaboration reflects her commitment to integrating practical exposure into academic research, thereby adding value to both the organization and the research ecosystem.

Her efforts have significantly supported the improvement of our recruitment policies and have provided a structured platform that benefits both the company and prospective candidates.

We appreciate her sincere engagement and wish her continued success in her academic and professional pursuits.

With best regards,

SHREE MAHAKALI SOFTWARE PRIVATE LIMITED

E-mail : info@sm-software.in

www.sm-software.in

Corporate Office :
1010 Fortune Business Hub,
Science City Road, Ahmedabad

Regd. Office :
A/12, Rivera Harmony, Survey No. - 392/1,
Opp. Royal Archad, Nr. Pinakal, Prahladnagar, Ahmedabad
CELL : 94277 45635

GSTIN : 24ABJCS7567F1ZZ | CIN : U72900GJ2022PTC136698

Appendix – II

Sathwara InfoTech

Date : 16-March-2025

To Whom It May Concern

We are pleased to acknowledge the contribution of **Ms. Monica Gupta** in the Campus Recruitment Drive conducted in collaboration with our organization.

The placement process model proposed by was well-structured, clear, and professionally executed. It helped improve communication between the company and candidates, made the recruitment process more efficient, and ensured smooth and transparent operations.

As part of this work, she successfully applied the recruitment parameters and evaluation criteria provided by our team. We are happy to note that her method delivered positive and measurable results, making the process more effective and outcome-focused.

We also understand that this work is part of her ongoing Ph.D. research at Gujarat Technological University (GTU). Her efforts reflect a strong commitment to connecting academic research with real-world industry practices.

Her contribution has helped improve our recruitment strategy while also providing a structured and practical process that benefits both the company and the candidates.

With Best Regards,
Mr. Bhavesh Sathwara
CEO And Founder.
Sathwara InfoTech

Sathwara InfoTech

+91-9104081291 | info@sathwarainfotech.com | www.sathwarainfotech.com
326, Arth Business Center (ABC), Opp. Torrent Power, Nikol Sp Ring Road, Ahmedabad, Gujarat 380038

Appendix – III

WEBOSPHERE

Date: 23-March-2025

To Whom It May Concern

This is to affirm the valuable contribution of **Ms. Monica Gupta** in designing and supporting the Campus Placement Activities undertaken in coordination with our organization.

The placement management methodology proposed by her reflected a strong sense of professionalism, strategic insight, and systematic structure. It effectively improved communication between the company and potential candidates, optimized the recruitment workflow, and ensured that the entire process was carried out in a seamless, efficient, and transparent manner.

Her proposed framework effectively incorporated our recruitment parameters and evaluation criteria, leading to measurable improvements and a more outcome-driven placement process.

This initiative forms part of her **ongoing Ph.D. research at Gujarat Technological University (GTU)** and highlights her commitment to bridging academic research with practical industry applications.

Her efforts have greatly improved our recruitment process, providing a structured approach that benefits both the company and prospective candidates. We sincerely appreciate her dedication and wish her continued success in her academic and professional pursuits.

With best regards,

WEBOSPHERE TECHNOLABS LLP LLPIN AAN-6015

Bharti Pareek

HR Executive
Webosphere Technolabs LLP