STUDENT PLACEMENT PREDICTION

A Project Report submitted for the fulfillment of one of the requirements

For the award of the Degree of

BACHELOR OF TECHONOLOGY
In

CSE-ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by

N.Naveen

Regd.No.20811A4220

G.Madhuri

Regd.No.20811A4206

R.Chandrika

Regd.No.20811A4224

Under the guidance of

Mr. N. DURGA PRASAD M.Tech(Ph.D)

Associate Professor

Department of Computer Science and Engineering-AI&ML

AVANTHI INSTITUTE OF ENGINEERING & TECHNOLOGY

(Approved by AICTE, New Delhi &permanently affiliated to JNTU Gurajada Vizianagaram)
(Accredited by NAAC A+, UGC & NBA, AICTE)
MAKAVARAPALEM, NARSIPATNAM, VISAKHAPATNAM DIST.

(2020-2024)

AVANTHI INSTITUTE OF ENGINEERING & TECHNOLOGY

(Approved by AICTE, New Delhi & Permanently affiliated to JNTU Gurajada Vizianagaram)
(Accredited by NAAC A+, UGC & NBA, AICTE)

MAKAVARAPALEM, NARSIPATNAM,
VISAKHAPATNAM DIST (2020-2024)

CERTIFICATE

This is to certify that the project entitled, "STUDENT PLACEMENT PREDICTION" in partial fulfilment for the degree of Bachelor of Technology in CSE – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, at AVANTHI INSTITUTE OF ENGINEERING & TECHNOLOGY, MAKAVARAPALEM, VISAKHAPATNAM is an bonafide work carried out by N.NAVEEN(20811A4220), G.MADHURI(20811A4206), R.CHANDRIKA(20811A4224) under the guidance N.DURGA PRASAD M.Tech(Ph.D) and supervision during 2023-2024.

Internal Guide
Mr. N. DURGA PRASAD M.Tech(Ph.D)
Associate professor
Department of CSE

Head of the Department
N.V.ASHOK KUMAR M.Tech,Ph.D
Associate Professor
Department of CSE-AI&ML

External Examiner

ACKNOWLEDGEMENT

We are thankful to N. DURGA PRASAD M.Tech (Ph.D) ASSOCIATE PROFESSOR, AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY for guiding us through this project giving his valuable suggestions and helping us to overcome the difficulties faced during design and coding stages of our project.

We are grateful to Head of the Department, N .V. Ashok Kumar M. Tech Ph.D Associate Professor, CSE-AI&ML, for providing us the required software and hardware encouraging us right through the project.

Our sincere thanks to Dr C.P.V.N.J. MOHAN RAO, M. TECH, PhD, Principal of AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY for being a source inspiration and constantly encouraging us throughout the course to pursue new goals and ideas.

We would like to thank our College Management for providing various resources to complete our project work successfully.

We are also thankful to my friends for the cooperation in our work and grateful thanks to my parents for giving us moral support in this work.

SUBMITTED BY

N.NAVEEN-20811A4220

G.MADHURI-20811A4206

R.CHANDRIKA-20811A4224

DECLARATION

We are hereby declare that the project entitled “STUDENT PERFORMANCE PREDICTION” is a bonafied work done by us and submitted for the partial fulfilment of the requirements for the award of degree of Bachelors of Technology in Computer Science & Technology – AIML from Jawaharlal Nehru Technological University - Gurajada Vizianagaram& approved by AICTE is my original work in the year 2023 – 2024 under the esteemed guidance of Mr. N. DURGA PRASAD ,M.Tech(PhD),ASSOSIATE PROFESSOR in the stream of computer science and engineering department and it is not previously formed. The basis for any degree or diploma or any other similar title submitted to any university.

PROJECT ASSOCIATES

N. NAVEEN - 20811A4220

G.MADHURI-20811A4206

R.CHANDRIKA-20811A4224

Date:

Place:

Abstract

This project develops a student placement prediction model utilizing logistic regression, designed to forecast the likelihood of high school students being placed in job roles based on academic records and other relevant factors. Leveraging data on students' performance in secondary education, including percentages in 10th and 12th grades, engineering degree scores, and quantitative and verbal abilities, the model aims to provide insights into placement outcomes. Additionally, attributes such as gender, type of education board, and extracurricular participation are considered. The model is deployed using Streamlit for interactive use, allowing users to input their details and receive predictions regarding their placement prospects. This research underscores the potential of machine learning techniques in guiding career decisions and enhancing educational outcomes.Beyond academic achievements, the model accounts for various influential factors, including gender, the type of education board attended, and participation in extracurricular activities. These additional attributes contribute to a comprehensive understanding of students' backgrounds and capabilities, enhancing the model's predictive accuracy.This model helps the position cell at intervals a corporation to spot the potential students and concentrate to and improve their technical and social skills.

TABLE OF CONTENTS

LIST OF CONTENTS

Chapter1

INTRODUCTION

1.1 OVERVIEW

We aim to develop a placement predictor as a part of making a placement management system at college level which predicts the probability of students getting placed. It will also help the teachers as well as placement cell in an institution to provide proper care towards the improvement of students in the duration of course. We are using machine learning for the placement prediction. The existing placement prediction model considers only academic performances of the students so that the prediction of the student getting placed or not can be done. We cannot consider the placement of students just by their academic performances because some students may be good at aptitude, technical and communication skills due to their low score in their academic that may tend to be their drawback. For predicting the placement of a Student needs parameters like cgpa, logical and technical skills Academic performances may be important but the model is design to predict the placements based on the parameters of the students and segregating the total students placed data according to their streams to find out that in which stream placements are more. Using this model we can predict the next year admission trends.

1.2 PROBLEM STATEMENT

The problem statement for student placement prediction is to determine the likelihood of a student getting placed in a job after completing their education.

The objective is to develop a model or algorithm that can accurately predict whether a student will get placed or not based on various factors like academic performance, skills, internships, and other relevant factors. This can help educational institutions and students make informed decisions and take necessary steps to improve their chances of getting placed.

1.3 EXISTING SYSTEM

The current system generally uses only a single parameter to judge whether a student can be placed or not during the campus placements. Generally the parameter used to judge the strengths of the student, is the academic performances during the first three years of engineering.

But cracking an interview not only depends on the academic scores but also the awareness of student during the aptitude tests and interviews. Also some Data Mining algorithms, while calculating the probability of a student getting selected, sometimes interpret the result having a probability of more than 100% which is not feasible and denotes a wrong interpretation to the student.

Some algorithms give a negative probability which gives an wrong interpretation to the student. Judging the student only on the basis of academic grades is not enough. The other parameters like aptitude and technical tests should also be taken into consideration in order to determine the outcome for the students future.

1.4 PROPOSED SYSTEM

In proposed system, we are not considering the placement of students just by their academic perform ances because some students may be good at aptitude, technical and communication skills due to thei r low score in their academic that may tend to be their drawback. For predicting the placement of a St udent needs parameters like cgpa, logical and technical skills Academic performances may be import ant but the model is design to predict the placements based on the parameters of the students and segr egating the total students placed data according to their streams to find out that in which stream place ments are more. Using this model we can predict the next year admission trends.

ADVANTAGES:

1. Proposed model can predict the placement more accurately because of we are using student's technical skills, communication skills, and also the internships done by the students in this predictive model.
2. We can use this model to predict the next year admission trends

1.5 Applications of Student Placement Prediction:

- Personalized Career Guidance: Placement predictions, combined with academic records and skills assessments, can help students identify strengths and weaknesses. This empowers them to make informed decisions about majors, internships, skill development, and ultimately, career paths.
- Increased Job Market Confidence: Knowing their predicted placement potential allows students to approach job applications and interviews with greater confidence. They can tailor their resumes and highlight relevant skills based on the specific job requirements.

- Early Intervention and Support: Students at risk of lower placement chances can be flagged early. This allows institutions to offer targeted support programs like career counseling, skill development workshops, or internship opportunities to improve their competitiveness.
- Matching with Suitable Employers: Placement predictions can connect students with companies that align with their skills and projected placement outcomes. This can lead to a more efficient job search process and a higher chance of landing a fulfilling position.

For Educational Institutions:

- Improved Career Services: Institutions can leverage placement predictions to tailor career services and resources to individual student needs. This could involve targeted workshops, internship opportunities, or mentorship programs aligned with predicted career paths.
- Enhanced Employer Relations: Accurate placement data can demonstrate an institution's effectiveness in preparing students for the workforce. This attracts top employers for recruitment events and internship programs, creating a mutually beneficial ecosystem.
- Curriculum Development: Analyzing placement data and employer needs allows institutions to identify skill gaps and adapt their curriculum. This ensures graduates possess the skills and knowledge demanded by the job market.
- Program Evaluation and Improvement: Placement prediction models can be used to evaluate the effectiveness of different academic programs and identify areas for improvement. This allows institutions to continuously enhance the quality of education they provide.

Beyond Students and Institutions:

- Corporate Recruitment: Companies can leverage placement prediction data to pre-screen job applicants. This identifies candidates with a higher predicted chance of success, leading to a more efficient recruitment process. Additionally, analyzing placement data can help them identify skills lacking in graduates, allowing them to tailor internship programs and collaborate with universities to bridge those gaps.
- Government Policy and Workforce Development: Government agencies can analyze placement prediction data across institutions to identify emerging skills shortages in the workforce. This information can be used to develop targeted training programs and upskilling initiatives to address these gaps. Additionally, placement data can be used to evaluate the

effectiveness of government-funded educational programs in preparing students for the job market.

- Evaluate the effectiveness of government-funded educational programs in preparing students for the job market.

CHAPTER 2

LITERATURE REVIEW

S.No	Title and Author Names	Description	Algorithms and Technologies used
1	**"Prediction Model for Students Future Development by Deep Learning and TensorFlow Artificial Intelligence Engine" -** Wilton W.T. FOK, Y.S. He, H.H Au Yeung and K.Y. Law	A study to predict suitable course for the students, based on their behavior using Neural Network Technique. Tensor Flow engine includes number of intermediate node and number of deep learning layers are adjusted and compared.	➢ Deep Neural Networks ➢ Tensorflow
2	**"Prediction of Student Enrolment Using Data Mining Techniques" -** H. Sabnani, M. More, P. Kudale, S. Janrao,	This paper proposes the Prediction of Student Enrolment Using Data Mining Techniques. They have used the Apriori technique to analyze the behavior of students who are seeking admission to a particular college. They have also used the Naïve Bayes algorithm which will help students to choose the course and help them in the admission procedure. In their project, they were conducting a test for students who were seeking admissions and then based on their performance, they were suggesting students a course branch using Naïve Bayes Algorithm	➢ Apriori technique ➢ Naïve Bayes Algorithm
3	**"A Placement Prediction System Using K-Nearest Neighbors Classifier"-**	They have proposed a system to predict the probability of a student being placed using k-nearest	➢ K-nearest neighbour

	Animesh,G.,M,Vignesh.,Bysani, P.,Naini,D.	neighbour classifier. Euclidean Distance is used as a distance measure. The dataset involves academic history of a student such as 10th percentage, 12th percentage, aggregate of engineering scores and number of backlogs. Further post processing is done using scores of Technical skills, Communication skills, Analytical Skills and as well as Teamwork scores are used.	
4	**"Application of Data mining in predicting student placement" -** Karan.P,Prateek.B	They have presented a paper on com parative study of different data mini ng techniques such as J48 decision t ree algorithm, Naive Bayes for pred icting a student's placement in diffe rent types of companies. The attribu tes of the dataset include cumulative grade point, different subject scores and the type of company. Missing d ata values were replaced with mean value for numeric data and mode va lue for nominal data.	➢ Naive Bayes ➢ Decision tree

5	**"Data Mining Techniques for Campus Placement Prediction in Higher Education" -** Patel, T., Tamrakar	In this paper, the creator had driven evaluations on the utilization of info rmation digging frameworks for gro unds position supposition and use of WEKA programming for plan and e xecution. Different cut-off points wh ich could be considered for figuring understudy execution are the schola stic show, social limits, specific limi ts, capable plan and tries. Different pressing assessments like clear k-me an, Farthest-first convergence, segre gated assembling, moderate groupin g were used for model turn of event s. It was seen that the time is taken f or building clear k-mean, Farthest-fi rst intersection point and bound gath ering was just 0.02sec in regards to various evened out squeezing (0.09 sec) and thickness based collecting (0.08 sec).	➢ K- Means
6	**"Placement Prediction System using Logistic Regression"** - Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar,PPS	They presented the development of placement predictor system (PPS) u sing logistic regression model. They used Machine learning technique to design and implement a logistic cla ssifier that predicts the probability o f the student to get placed along wit h Gradient Descent algorithm.	➢ Logistic Regression

7	"**An Empirical Analysis of Classification Techniques for Predicting Academic Performance"** - S.Taruna , Mrinal Pandey	They implemented an empirical analysis on predicting academic performance by using classification techniques or mapping of data items into predefined groups and classes using supervised learning. They compared five classification algorithms namely Decision Tree, Naive Bayes, Naive Bayes Tree, K-Nearest Neighbour and Bayesian Network algorithms for predicting student's grade particularly for engineering students using a four class prediction problem.	➢ Decision Tree, ➢ Naive Bayes, ➢ K-Nearest Neighbour ➢ Bayesian Network algorithms
8	**"Predicting students marks in hellenic open university" -** Kotsiantis, Sotiris B., and Panayiotis E. Pintelas,	They predicted the student marks (pass and fail classes) using the regression methods and available previous data. A number of experiments have been conducted with six algorithms, which were trained using datasets provided by the Hellenic Open University.	➢ Regression
9	**"Improving accuracy of students final grade prediction model " -** Syed Tanveer Jishan, Raisul Islam Rashu, Naheena Haque and Rashedur M Rahman	They proposed a model to predict student performance considering more on academic records. They accuired higher by applying some algorithms like Naive Bayes, Decision Tree and Neural Network	➢ Naive Bayes, ➢ Decision Tree ➢ Neural Network

10	**"ID3 Decision Tree Algorithm for Placement Prediction**" - Hitarthi Bhatt, Shraddha Mehta and Lynette R. D'mello	They identified relevant attributes based on quantitative and qualitative aspects of a student's profile such as CGPA, academic performance, technical and communication skills and designed a model which can predict the placement of a student	➢ ID3 Decision Tree Algorithm
11	**" Prediction and Analysis for Students Marks Based on Decision Tree Algorithm" -** Zhiwu Liu and Xiuzhi Zhang	They used decision tree algorithm to establish a classification rule and an analysis- forecasting model for students marks	➢ Decision Tree

CHAPTER 3

PROPOSED METHODOLGY

ALGORITHMS

LOGISTIC REGRESSION ALGORITHM

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

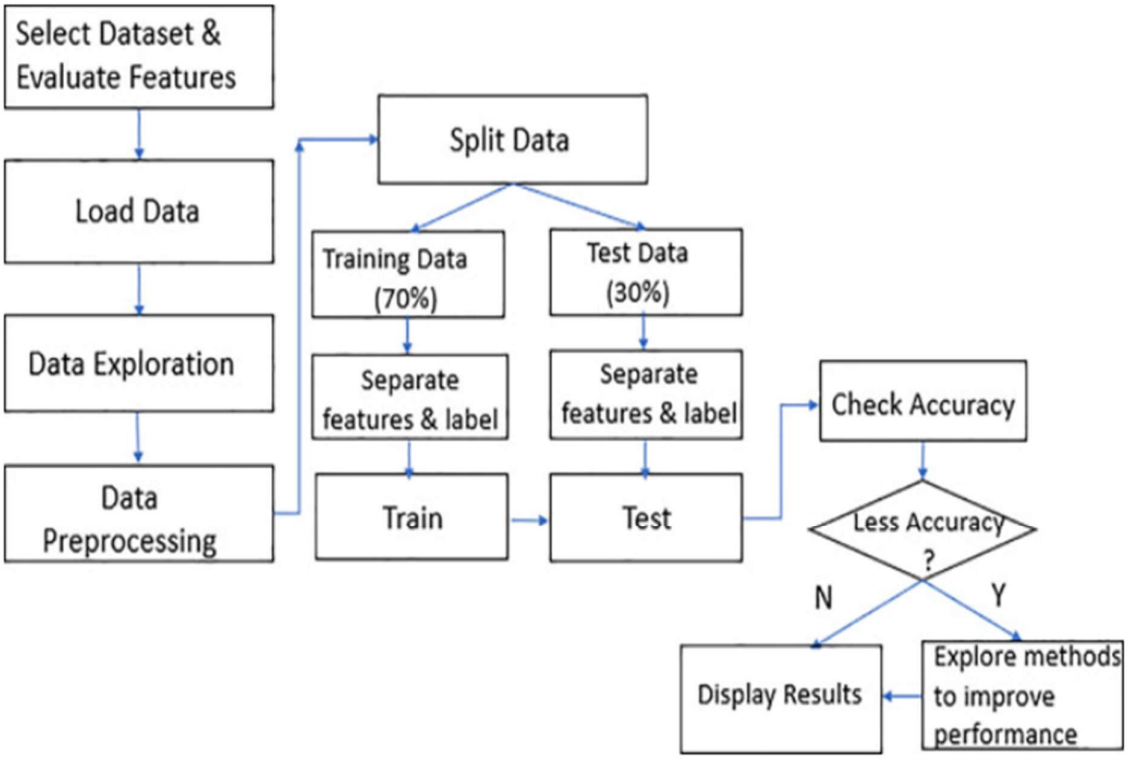


Figure.1

HOW DOES LOGISTIC REGRESSION WORK?

The Logistic regression working can be explained on the basis of the below algorithm

logistic regression works in terms of the algorithm:

1. Data Preparation Logistic regression requires labeled data for training. Each observation in the dataset should have one or more features (independent variables) and a binary outcome (dependent variable) indicating the class membership (e.g., 0 or 1, True or False).

2. Model Initialization In logistic regression, we initialize the model with coefficients (weights) and an intercept term. Initially, these coefficients can be set randomly or to zero.

3. Hypothesis Function Logistic regression uses a hypothesis function that maps the input features to the output labels. The hypothesis function is a sigmoid function, also known as the logistic function, which ensures that the output of the function lies between 0 and 1. The hypothesis function is defined as:

\[h_\theta(x) = \frac{1}{1 + e^{-(\theta^Tx + b)}} \]

where:

- (h_\theta(x) \) is the predicted probability that \(y = 1 \) given input \(x \).

- (theta) is the vector of coefficients (weights).

- (x) is the input feature vector.

- (b) is the intercept term.

4. Cost Function Logistic regression uses the logistic loss function (also known as the cross-entropy loss) to measure the error between the predicted probabilities and the actual labels. The cost function is defined as:

[J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))] \]

where:

- \(m \) is the number of training examples.

- \(y^{(i)} \) is the actual label for the \(i \)-th example.

- \(h_\theta(x^{(i)}) \) is the predicted probability for the \(i \)-th example.

5. Gradient Descent: To minimize the cost function, we use optimization algorithms like gradient descent. Gradient descent iteratively updates the coefficients and intercept term by moving in the direction of steepest descent of the cost function. The update rule for gradient descent is:

\[\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \]

where:

- \(\alpha \) is the learning rate, which controls the step size of the updates.

6. Model Training: We repeat the gradient descent process until the cost function converges to a minimum or until a predefined number of iterations is reached. This process optimizes the coefficients and intercept term of the logistic regression model.

7. Prediction: Once the model is trained, we can use it to make predictions on new data by applying the hypothesis function to the input features.

3.1 SYSTEM REQUIREMENTS

System requirements are a statement that identifies the functionality that is needed by a system in order to satisfy the customer's requirements. System requirements are a broad and also narrow subject that could be implemented to many items. Whether discussing the system requirements for certain computers, software, or business processes from a broad viewpoint. Also, taking it down to the exact hardware or coding that runs the software. System requirements are the most effective way of meeting user needs and reducing the cost of implementation.

System requirements are categorized into two categories, that is hardware and software requirements.

Hardware requirements:

- Processing power: A computer with sufficient processing power (CPU) is needed to handle data processing, model training, and potentially user requests. This might be a single powerful machine or a distributed computing system depending on data volume and model complexity.
- Memory (RAM): Adequate RAM is crucial for handling data loading, manipulation, and model operations during training and prediction.
- Storage: Sufficient storage space is required to store the collected student data, potentially anonymized, and the trained model itself.

Software requirements:

- Operating System: A stable operating system (OS) like Windows, Linux, or macOS is needed to run the system software.
- Programming Languages: Languages like Python

Data requirements:

- Student Data: This is the core element for training the model. The data should encompass factors potentially influencing placement success, such as academic performance (GPA, transcripts), standardized test scores, skills assessments, internship experiences, extracurricular activities, and potentially anonymized demographic information.
- Data Security and Privacy: Ensure mechanisms are in place to protect student data privacy throughout the system lifecycle, following relevant data privacy regulations.

3.2 SYSTEM ARCHITECTURE

System architectureis used for predicting student placement success using machine learning techniques.

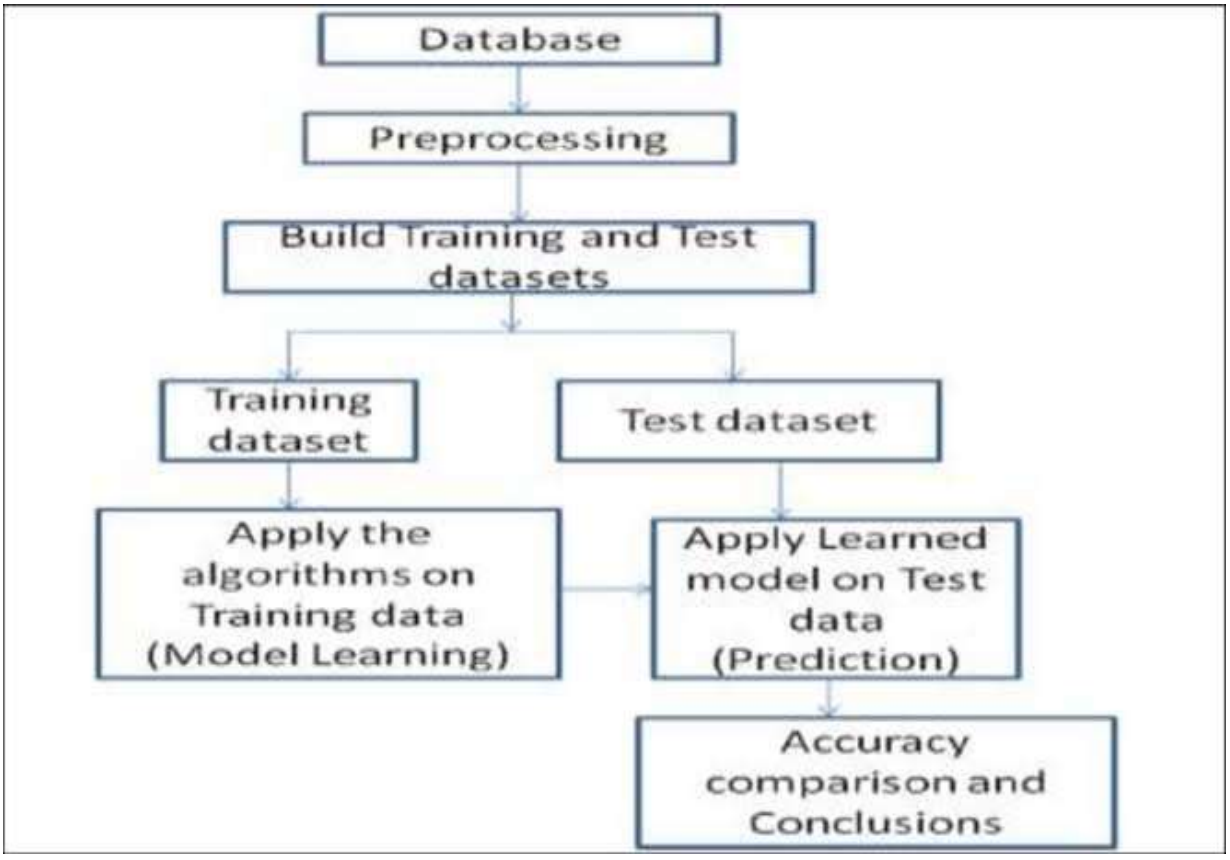


Figure.2

System Goal:

The goal of this system is to develop a model that can accurately predict the likelihood of a student securing a job placement after graduation. This information can be valuable for students, advisors,

and educational institutions in making informed decisions about career paths and educational programs.

System Design:

1. **Data Acquisition:**

    - Gather data on students from various sources:
    - Academic records (GPA, transcripts)
    - Standardized test scores (if applicable)
    - Skills assessments (technical and soft skills)
    - Internship experiences (projects, duration)
    - Extracurricular activities (leadership roles)
    - Demographic information (optional, anonymized)
    - Company placement requirements (if possible)
    - Ensure data privacy and anonymize sensitive information.

2. **Data Preprocessing:**

    - Clean the data by addressing missing values, inconsistencies, and outliers.
    - Encode categorical variables (e.g., major) into numerical representations for machine learning algorithms.
    - Feature engineering: Create new features that might be more relevant for prediction. Examples include:
    - A binary feature for having a GPA above a certain threshold.
    - Combining scores from related skills tests into a single feature.

3. **Model Building and Training:**

    Logistic Regression: A good starting point due to its interpretability. However, it assumes linear relationships, which may not hold true for all factors.

4. **Model Evaluation:**

    - Evaluate the model's performance on the testing data using metrics like:
    - Accuracy: Overall percentage of correct predictions.

- Precision: Ratio of true positives to all predicted positives.
- Recall: Ratio of true positives to all actual positives.
- F1 Score: Harmonic mean of precision and recall.

5. **Model Deployment:**

- Deploy the model into a user-friendly system that allows:
- Students and advisors to input student data.
- The model to generate predictions on the likelihood of placement.
- Visualization of results and insights into factors that influence predictions (possible with interpretable models like logistic regression).

Methodology:

1. **Data Collection:** Collaborate with educational institutions and relevant organizations to gather student data ethically and securely.
2. **Data Exploration:** Analyze the collected data to understand the relationships between features and placement outcomes.
3. **Feature Selection:** Identify the most relevant features for model training based on data exploration and domain knowledge.
4. **Model Training and Evaluation:** Implement the chosen machine learning algorithm, train on the prepared data, and evaluate its performance on the testing set. Iterate on model selection, hyperparameter tuning, and feature engineering if needed.
5. **Model Deployment and Refinement:** Deploy the final model for user interaction and continuously monitor its performance over time. Gather new data and retrain the model periodically to maintain accuracy.

CHAPTER 4

4.DATA COLLECTION AND PREPROCESSING TECHNIQUES

4.1 IDENTIFYING RELEVANT DATA SOURCES

1. Academic Performance Data:

- **Institutional Data Warehouses:** These store student records including GPA, course grades, major, and standardized test scores (if applicable).
- **Learning Management Systems (LMS):** Data from LMS platforms can reveal student performance in online courses, completion rates, and engagement metrics.

2. Skill and Experience Data:

- **Student Resumes and Portfolios:** Allowing students to upload resumes and portfolios provides valuable insights into their skills, projects, and accomplishments.
- **Skill Assessment Platforms:** Third-party platforms offering standardized skill assessments in technical areas, soft skills, or specific industry knowledge can be integrated.
- **Internship and Project Records:** Data on internship experiences, projects undertaken (individual or group), and relevant certifications can be valuable indicators of skills and practical experience.

3. Career-Related Activities:

- **Career Center Data:** Information on workshops attended, career fairs participated in, and career counseling sessions can indicate student career interests and initiative.
- **Co-curricular Activities and Leadership Roles:** Data on involvement in student clubs, organizations, leadership positions held, and volunteer work showcases soft skills and extracurricular achievements.

4. Additional Considerations:

- **Third-Party Data Integration (with Consent):** Exploring partnerships with relevant platforms like online learning providers or internship portals can offer additional data on student skills and experiences (ensure student consent and data privacy are addressed).
- **Self-Reported Data:** Allowing students to self-report their interests, career goals, and relevant skills through surveys or questionnaires can provide valuable insights, though verification methods might be needed.

4.2 Data preprocessing techniques

1. Handling Missing Values:

- **Deletion:** If missing data points are minimal and unlikely to bias the results, you can remove them. However, this approach might be less favorable if you have a large dataset.
- **Mean/Median/Mode Imputation:** Replacing missing values with the average (mean), middle value (median), or most frequent value (mode) for that specific feature can be a simple solution. However, it might introduce bias if the missing values are not random.
- **Machine Learning Techniques:** Techniques like k-Nearest Neighbors (KNN) or Expectation-Maximization (EM) can be used to predict missing values based on similar data points.

2. Dealing with Outliers:

- **Capping:** Extreme outliers can be capped to a specific value within a reasonable range to prevent them from unduly influencing the model.
- **Winsorization:** Similar to capping, this technique replaces outliers with values at a certain percentile (e.g., 95th percentile) of the data distribution.
- **Removal:** If outliers are genuine errors or anomalies, they can be removed, especially if they represent a very small portion of the data.

3. Encoding Categorical Features:

- **Label Encoding:** Assigning a numerical value to each category (e.g., "Major: Computer Science" = 1, "Major: Mathematics" = 2). This is a simple approach, but it might introduce an

order that doesn't exist in the categories (e.g., assuming CS is "better" than Math because it has a higher value).

- **One-Hot Encoding:** Creating a new binary feature for each category (e.g., "Major_CS" = 1, "Major_CS" = 0, "Major_Math" = 0, "Major_Math" = 1). This avoids the ordinal bias issue but can lead to a high number of features if there are many categories.

4. Feature Scaling:

- **Normalization:** Scaling features to a common range (e.g., 0 to 1) can be beneficial when features have different units or scales (e.g., GPA vs. number of internship experiences).
- **Standardization:** Transforming features to have a mean of 0 and a standard deviation of 1. This is useful when the distribution of features is not necessarily uniform.

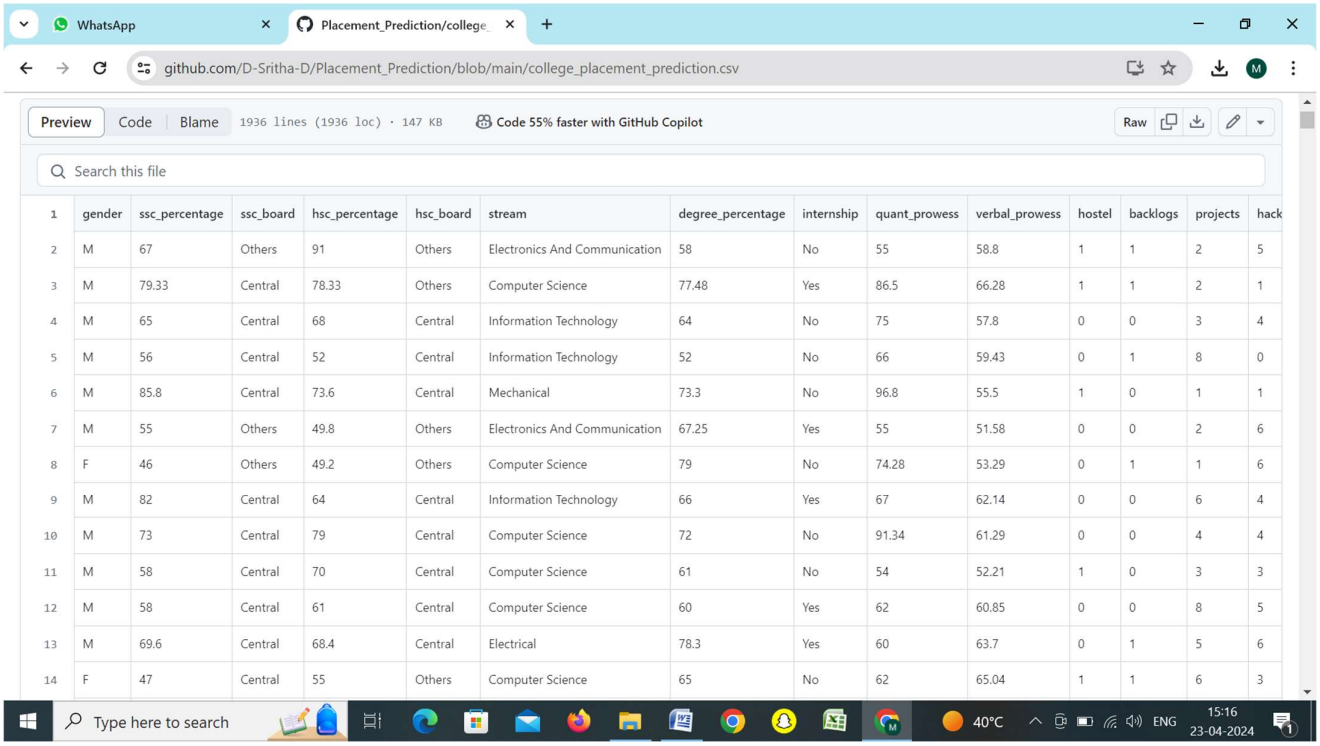
5. Feature Selection:

- **Correlation Analysis:** Identifying features that are highly correlated with each other might indicate redundancy. You can keep the most informative feature and remove the others.
- **Feature Importance Techniques:** Techniques like recursive feature elimination or feature importance scores from machine learning models can help identify the most relevant features for prediction.

4.3 Data collection methods

1. Institutional Data

- **Student Information Systems (SIS):** Extract data on student demographics (age, gender), academic records (GPA, course grades, major, standardized test scores), and disciplinary records (if applicable).
- **Learning Management Systems (LMS):** Gather information on student performance in online courses, completion rates, and engagement metrics (time spent on modules, attempts on quizzes).
- **Career Center Records:** Collect data on workshops attended, career fairs participated in, career counseling sessions utilized, and internship placements secured through the career center.



1	gender	ssc_percentage	ssc_board	hsc_percentage	hsc_board	stream	degree_percentage	internship	quant_prowess	verbal_prowess	hostel	backlogs	projects	hack
2	M	67	Others	91	Others	Electronics And Communication	58	No	55	58.8	1	1	2	5
3	M	79.33	Central	78.33	Others	Computer Science	77.48	Yes	86.5	66.28	1	1	2	1
4	M	65	Central	68	Central	Information Technology	64	No	75	57.8	0	0	3	4
5	M	56	Central	52	Central	Information Technology	52	No	66	59.43	0	1	8	0
6	M	85.8	Central	73.6	Central	Mechanical	73.3	No	96.8	55.5	1	0	1	1
7	M	55	Others	49.8	Others	Electronics And Communication	67.25	Yes	55	51.58	0	0	2	6
8	F	46	Others	49.2	Others	Computer Science	79	No	74.28	53.29	0	1	1	6
9	M	82	Central	64	Central	Information Technology	66	Yes	67	62.14	0	0	6	4
10	M	73	Central	79	Central	Computer Science	72	No	91.34	61.29	0	0	4	4
11	M	58	Central	70	Central	Computer Science	61	No	54	52.21	1	0	3	3
12	M	58	Central	61	Central	Computer Science	60	Yes	62	60.85	0	0	8	5
13	M	69.6	Central	68.4	Central	Electrical	78.3	Yes	60	63.7	0	1	5	6
14	F	47	Central	55	Others	Computer Science	65	No	62	65.04	1	1	6	3

Figure.3

2. Student Self-Reported Data

- **Surveys and Questionnaires:** Develop surveys to assess student career interests, desired job roles, relevant skills (technical and soft skills), extracurricular activities, leadership positions held, volunteer work, and self-reported GPA (if not accessible through SIS).
- **Skill Assessment Platforms:** Integrate with third-party platforms offering standardized skill assessments in technical areas relevant to target job roles (e.g., programming languages, data analysis) or soft skills assessments.
- **Resume and Portfolio Uploads:** Create a secure system for students to upload resumes and portfolios. This allows them to showcase their skills, projects, accomplishments, and relevant work experiences.

3. External Data Sources (with Consent)

- **Third-Party Learning Platforms:** Partner with online learning platforms students might use to access data on additional courses taken, certifications earned, and project completion details (ensure student consent and data privacy are addressed).

- **Internship Portals:** Collaborate with internship portals where students might find opportunities to access data on internship experiences not secured through the career center (again, with student consent and data privacy considerations).

Optimizing Data Collection for Logistic Regression:

- **Focus on Relevant Features:** Logistic regression works best with a well-defined set of features that are demonstrably linked to student placement success.
- **Feature Selection Techniques:** Employ techniques like correlation analysis or feature importance scores from initial model runs to identify the most relevant features for placement prediction.
- **Data Balance:** In cases where successful placements are less frequent, consider employing data balancing techniques (e.g., oversampling or undersampling) to ensure the model is not biased towards the majority class.

CHAPTER 5

5. FACTORS INFLUENCING STUDENT PLACEMENT

Success in student placement hinges on a combination of factors from a student's academic background, practical experiences, and personal qualities. Here's a breakdown of some key influences:

Academic Performance:

- **GPA:** A strong GPA demonstrates a student's ability to grasp complex concepts, manage workload, and excel in academic environments.
- **Coursework:** Relevant coursework directly tied to the target job roles showcases a student's foundational knowledge and specific skill sets.

Skills and Abilities:

- **Technical Skills:** Depending on the job role, proficiency in programming languages, data analysis tools, or industry-specific software can be crucial.
- **Soft Skills:** Communication, teamwork, problem-solving, critical thinking, and adaptability are essential skills sought by employers across various fields.

Other Factors:

- **Career Interests and Goals:** A student's genuine interest and passion for the targeted field can significantly influence their motivation and performance during placement.
- **Communication and Presentation Skills:** The ability to articulate ideas clearly, confidently, and professionally is essential for making a strong impression during interviews.
- **Network and References:** Building connections with professionals in the field can open doors to opportunities and provide valuable references that strengthen a student's application.

CHAPTER 6

6.FEATURE ENGINEERING

Feature engineering is a crucial step in machine learning, particularly for tasks like student placement prediction using logistic regression. It's the process of transforming raw data into features that are most effective for training a machine learning model.

6.1 Feature Selection

- **Correlation Analysis:** Analyze the correlation between features to identify redundant or highly correlated ones. Removing redundant features reduces model complexity and avoids multicollinearity, which can lead to unstable model coefficients.
- **Feature Importance Techniques:** Train a preliminary logistic regression model and analyze the feature importance scores. Focus on features with the highest scores, indicating a strong influence on the prediction of successful placements.
- **Domain Knowledge:** Leverage your understanding of the factors influencing placement success to select relevant features. For example, GPA and internship experience are likely more informative than extracurricular activities not directly related to the target job roles.

6.2 Feature Extraction Techniques:

- **Feature Creation:** Combine existing features to create new ones that might be more predictive. For instance, create a feature like "GPA x Internship Experience" to capture the combined effect of academic performance and practical experience.
- **Text Analysis (if applicable):** If student resumes or project descriptions are included, use techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to extract relevant keywords and topics, transforming textual data into numerical features suitable for logistic regression.

6.3 Feature Transformation Methods:

- **Normalization and Standardization:** Logistic regression performs better when features are on a similar scale. Apply normalization (scaling to a range like 0-1) or standardization

(transforming to have a mean of 0 and standard deviation of 1) to ensure all features contribute equally to the model.

- **Encoding Categorical Features:** Logistic regression requires numerical features. Encode categorical features like "Major" or "Job Role" using techniques like one-hot encoding (creating a binary feature for each category) or label encoding (assigning a numerical value to each category, but be cautious of introducing ordinal bias).

CHAPTER-7

7. MODEL SELECTION

7.1. Overview of Predictive Modelling Techniques:

Logistic Regression: A linear model well-suited for binary classification problems (placed vs. not placed). It's interpretable, making it easier to understand how features influence predictions. However, it might struggle with complex non-linear relationships between features.

7.2. Evaluation Criteria for Model Selection:

- **Accuracy:** The percentage of predictions that are correct (correctly classified placements and non-placements).
- **Precision:** The proportion of predicted placements that actually turn out to be placements.
- **Recall:** The proportion of actual placements that are correctly predicted by the model. (Also known as True Positive Rate)

7.3. Choosing the Appropriate Model for Student Placement Prediction:

For Interpretability: If understanding how features influence predictions is crucial, logistic regression or decision trees might be preferable.

For Accuracy and Non-linear Data: If accuracy is paramount, and you suspect non-linear relationships in your data, consider random forests or support vector machines.

For Complex Relationships and Large Data: When dealing with very complex relationships and a vast amount of data, deep learning models might be an option, but interpretability becomes a challenge.

Model Training Process:

1. **Preprocessing:** Apply data cleaning and preprocessing techniques you defined earlier to ensure your data is ready for training.

2. **Model Training:** Feed the training data to your chosen model. The model learns the relationships between features and the target variable (placement vs. not placed) by adjusting its internal parameters to minimize prediction errors.
3. **Model Fitting:** The training process essentially "fits" the model to the training data. This involves finding the optimal parameters that allow the model to make accurate predictions.

CHAPTER 8

8. MODEL TRAINING AND EVALUATION FOR STUDENT PLACEMENT PREDICTION

8.1 Data Splitting for Training and Testing

- **Random Split:** Divide your student placement data into two sets:
- **Training Set (70-80%):** This larger portion is used to train the logistic regression model. The model learns the relationships between features (GPA, internship experience, etc.) and the target variable (placed/not placed) by adjusting its internal parameters to minimize prediction errors.
- **Testing Set (20-30%):** This unseen data is used to evaluate the model's performance on new, real-world scenarios. By using unseen data, you get a more accurate picture of how well the model will generalize to future student placement predictions.

8.2 Model Training Process

- **Data Preprocessing (Recap):** Apply the cleaning and transformation techniques you defined earlier to ensure your data is ready for training (handling missing values, dealing with outliers, encoding categorical features, feature scaling, etc.).
- **Splitting the Data:** Divide your data as described in section 8.1, using random or stratified splitting as appropriate.
- **Model Training:** Feed the training data to the logistic regression model. The model performs a series of calculations to learn the relationships between features and the target variable. It iteratively adjusts its internal parameters (coefficients) to minimize prediction errors on the training data.
- **Model Fitting:** The training process essentially "fits" the model to the training data. This involves finding the optimal set of coefficients that allow the model to make the most accurate predictions possible based on the training data.

8.3 Performance Evaluation Metrics

- **Accuracy:** The percentage of correctly classified placements and non-placements (correctly predicted "placed" and "not placed" outcomes). A high accuracy is desirable, but it might not be the most informative metric in cases of imbalanced data.
- **Precision:** The proportion of predicted placements that actually turn out to be placements. This is crucial if incorrectly classifying a non-placement as a placement has significant consequences (e.g., wasted resources on an unsuitable candidate).
- **Recall (True Positive Rate):** The proportion of actual placements that are correctly predicted by the model. This is important if missing true placements is a major concern (e.g., overlooking a qualified student).

CHAPTER 9

OVERVIEW OF TECHNOLOGIES

Student placement prediction uses a combination of data analysis and machine learning techniques to forecast the likelihood of a student securing a job during campus recruitment. Here's a look at the core technologies involved:

1. Machine Learning Algorithms:

Logistic Regression: This is the workhorse of this project. It's a classification algorithm that analyzes student data (grades, skills, experience) and predicts the binary outcome of "placed" or "not placed" based on a calculated probability.

2. Data Management:

Python Libraries (Pandas, NumPy): These libraries are essential for handling student data. Pandas helps manipulate and structure the data (like organizing grades and skills), while NumPy provides tools for numerical computations needed for machine learning.

3. Machine Learning Frameworks:

Scikit-learn: This is a popular Python library specifically designed for building and deploying machine learning models. It provides all the functionalities needed for Logistic Regression in this project, including training the model, evaluating its performance, and making predictions on new student data.

Data Visualization (Matplotlib, Seaborn): These libraries can be used to explore and understand the student data better. You can visualize how different attributes like grades or skills correlate with placement success.

Web Development (Flask, Django): If you want to create a user-friendly application where students can input their details and get placement predictions, these frameworks can help build the interface and integrate the prediction model behind the scenes.

Overall Workflow:

- **Data Collection:** Gather historical placement data from your institution. This data should include student attributes and their placement outcomes (placed/not placed).
- **Data Wrangling:** Clean and prepare the data using Pandas and NumPy. This might involve handling missing values, converting categorical data (like branch of study) into numerical formats, and potentially scaling numerical features (like GPA) for better model performance.
- **Model Building:** Use Scikit-learn to train the Logistic Regression model on the prepared data. The model learns the relationships between student attributes and placement success.
- **Model Assessment:** Evaluate the model's performance on unseen data to ensure its accuracy and ability to generalize to new student profiles. Metrics like accuracy, precision, and recall help assess how well the model performs.
- **Prediction:** Once satisfied with the model, use it to predict placement probabilities for new students based on their attributes.

CHAPTER 10

IMPLEMENTATION AND CODING

Importing Libraries

```
Importnumpyas np
import pandas as pd
```

Reading the data

```
df = pd.read_csv("/content/college_placement_prediction.csv")
df.head()
```

Data Frame Information

```
df.info()
```

To know the gender

```
df['gender'].value_counts()
```

To replace the Gender with numerical values

```
df['gender'] = df['gender'].replace({'M': 1, 'F': 0}).astype(float)
```

to replace the ssc board with numerical values

```
df['ssc_board'] = df['ssc_board'].replace({'Others': 1, 'Central': 0}).astype(float)
```

To replace the HSC board with numerical values

```
df['hsc_board'] = df['hsc_board'].replace({'Others': 1, 'Central': 0}).astype(float)
```

To replace the status with numerical values

```
df['status'] = df['status'].replace({'Not Placed': 1, 'Placed': 0}).astype(float)
```

Finding correlation between the attributes

```
corr_df = df.drop(['gender', 'hostel', 'stream', 'internship' ,'backlogs', 'projects', 'hackathons',
'college_tier'], axis = 1)
```

To display the contents of the DataFrameof correlation matrix

```
print(corr_df)
```

Finding correlation coefficients between each pair of columns in the DataFrame

```
.corr = corr_df.corr()
```

To generate a heatmap visualizing the correlation matrix

```
import seaborn assns
sns.heatmap(corr,cmap='plasma',annot=True,fmt=".2f")
```

To create histograms of the 'hsc_percentage' and 'ssc_percentage' columns from your DataFrame

```
importmatplotlib.pyplotasplt
plt.hist(df['hsc_percentage'],bins=20)
plt.title("Distribution of HSC Percentage")
plt.xlabel('Percentage')
plt.ylabel('count')
plt.show()
importmatplotlib.pyplotasplt
plt.hist(df['ssc_percentage'],bins=20)
plt.title("Distribution of SSC Percentage")
plt.xlabel('Percentage')
plt.ylabel('count')
plt.show()
```

To create a scatter plot to visualize the relationship between 'ssc_percentage' and 'hsc_percentage' columns from your DataFrame

```
sns.scatterplot(x='ssc_percentage',y='hsc_percentage',data=df,hue='status')
plt.title("Correlation bw SSC and HSC Percentage")
```

```
plt.show()
```

function to convert categorical variables into dummy/indicator variables in your DataFrame

```
df=pd.get_dummies(corr_df,columns=[ 'ssc_board', 'hsc_board', 'status'],drop_first=True)
df
```

contains all the feature columns except 'status_1.0', and **y** contains the 'status_1.0'

```
columnx = df.drop('status_1.0',axis=1)
y = df['status_1.0']
```

print the shapes of your feature matrix **x** and target vector **y**

```
print(x.shape,'\n\n',y.shape)
```

To split data into training and testing sets for machinelearning

```
.fromsklearn.model_selectionimporttrain_test_split
```

to split your data into training and testing sets. Here's what your code does:

```
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=2)
```

```
print(X_train.shape,'\n\n',X_test.shape)
```

to access the testing set of your features after splitting your data

```
X_test
```

scikit-learn to train a logistic regression model on your training data

```
fromsklearn.linear_modelimportLogisticRegression
lg = LogisticRegression()
lg.fit(X_train,y_train)
y_pred = lg.predict(X_test)
```

To evaluate the accuracy of your logistic regression model

```
fromsklearn.metricsimportaccuracy_score
print(accuracy_score(y_test,y_pred))
```

This imports the **pickle** module, which is used for serializing and deserializing Python objects.

```
import pickle
pickle.dump(lg,open('/content/stdnt.pp','wb'))
```

The CSS code will be written to a file named **style.css**

```
%%writefile style.css
/ style.css/
body {
   background-color:#f0f2f6;
   font-family:Arial, sans-serif;
}

.widget.stButton>button {
   background-color:#4CAF50;
   color:white;
   padding:10px20px;
   text-align:center;
   text-decoration:none;
   display:inline-block;
   font-size:16px;
   margin:4px2px;
   transition-duration:0.4s;
   cursor:pointer;
}

.widget.stButton>button:hover {
   background-color:#45a049;
}
```

creating a web application using Streamlit and Python to predict job placement based on various input features

```
.%%writefile app.py
importnumpyas np
import pandas as pd
from PIL import Image
importstreamlitasst
import pickle

# Load CSS for styling
deflocal_css(file_name):
   withopen(file_name) as f:
      st.markdown(f'<style>{f.read()}</style>', unsafe_allow_html=True)
```

```
local_css("style.css")

# web app
st.title("Job Placement Prediction Model")

st.write('On entering the following Information, we will predict whether you will get placed or not')

# Inputs
name = st.text_input("Enter your name", "")
age = st.slider("Select your Age", 0, 100, 25)

gender_input = st.selectbox("Select your gender", ["Female", "Male"])
gender_M = 0ifgender_input == "Female"else1

ssc_percentage = st.text_input("Enter your 10th Class Percentage", "")
SSC_board_input = st.selectbox("Select your SSC Board", ["Central", "Other"])
ssc_board_Others = 0ifSSC_board_input == "Central"else1

hsc_percentage = st.text_input("Enter your 12th Class Percentage", "")
HSC_board_input = st.selectbox("Select your HSC Board", ["Central Board", "Other Board"])
hsc_board_Others = 0ifHSC_board_input == "Central Board"else1

options = ['Computer Science', 'Electronics and Communication', 'Information Technology',
'Mechanical', 'Civil', 'Electrical']
selected_option = st.selectbox('Select your Branch of Study for Engineering', options)

degree_percentage = st.text_input("Enter your Engineering Percentage", "")
quant_prowess = st.slider("Rate your Quantitative Ability (1-100)", 1, 100, 25)
verbal_prowess = st.slider("Rate your Verbal Ability (1-100)", 1, 100, 25)
internships = st.text_input("Number of internships", "")
backlogs = st.text_input("Number of backlogs", "")
projects = st.text_input("Number of projects", "")
hackathons = st.text_input("Number of hackathons participated", "")
college_tier = st.selectbox("Select your College Tier", [1, 2, 3])

# Load model
withopen('placement.pkl', 'rb') as f:
    model = pickle.load(f)

# Submit button
ifst.button("Submit"):
    # Make prediction
    predictions = model.predict([[float(ssc_percentage), float(hsc_percentage),
float(degree_percentage), float(quant_prowess), float(verbal_prowess), int(gender_M),
int(ssc_board_Others), int(hsc_board_Others)]])
    result = "Congratulations! You will be placed!"if predictions == 1else"Sorry, you will not be
placed. Keep working harder!"

    # Display result
```

```
   st.markdown(f"<h2 style='text-align: center; color: {'green'if predictions ==
1else'red'};'>{result}</h2>", unsafe_allow_html=True)
```

Installing the Streamlit library using pip

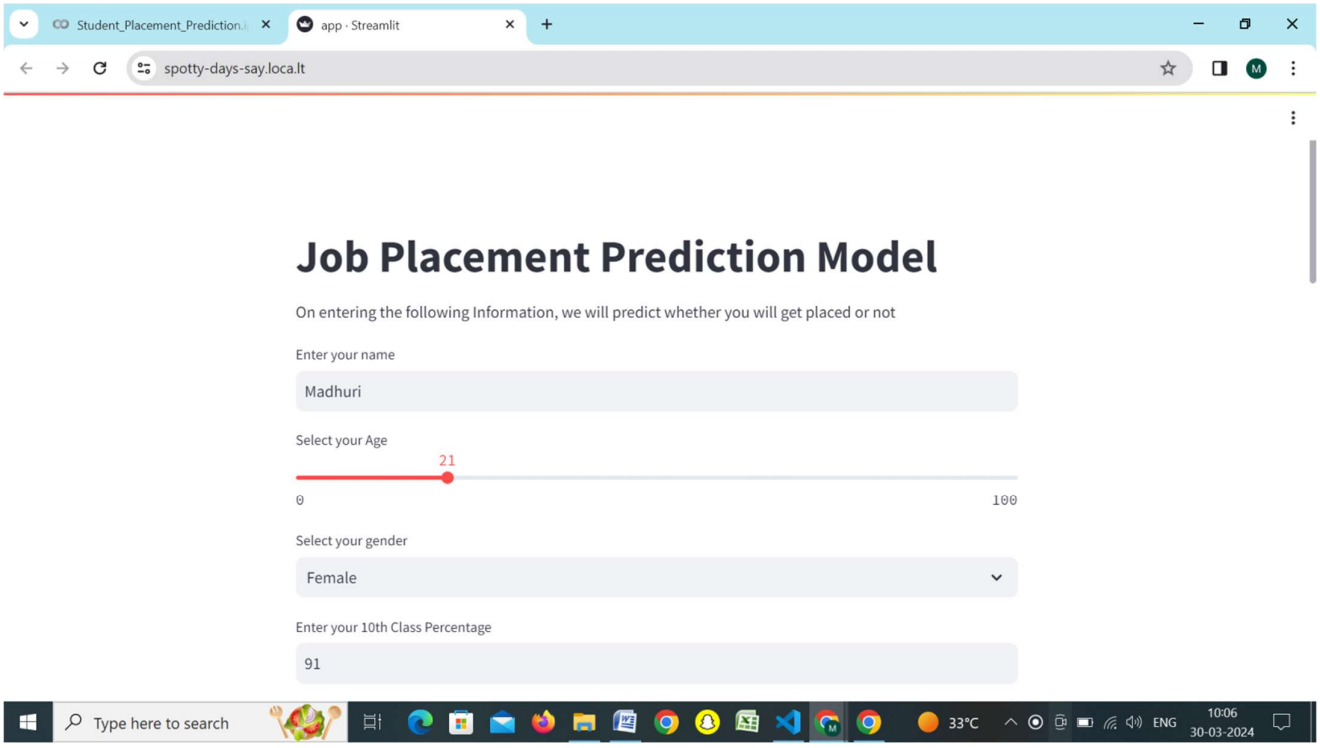
```
!pip install Streamlit –q
```

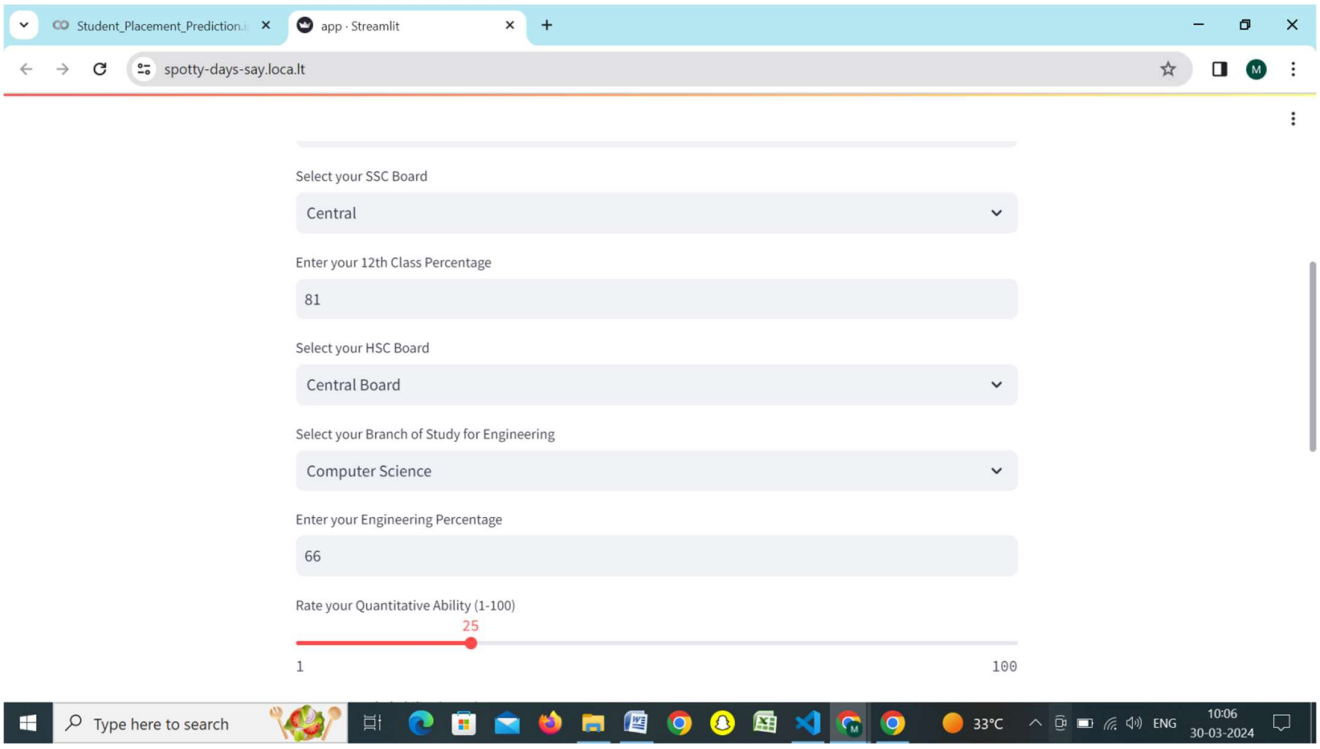
The code successfully retrieves the public IP address using the ipify API and prints it to the console.import requests

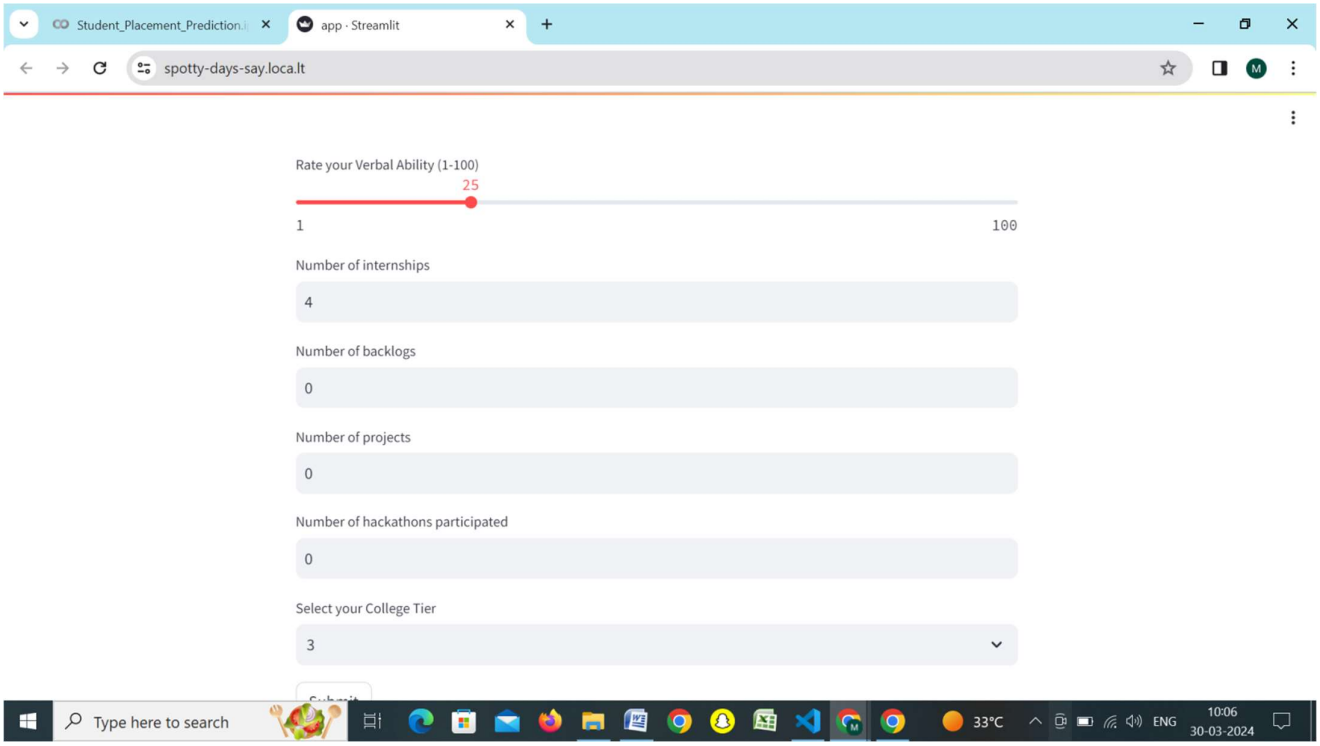
```
import requests

defget_public_ip():
   response = requests.get('https://api.ipify.org')
   return response.text

public  ip = get  public  ip()
print(f"Public IP Address: {public_ip}")
```

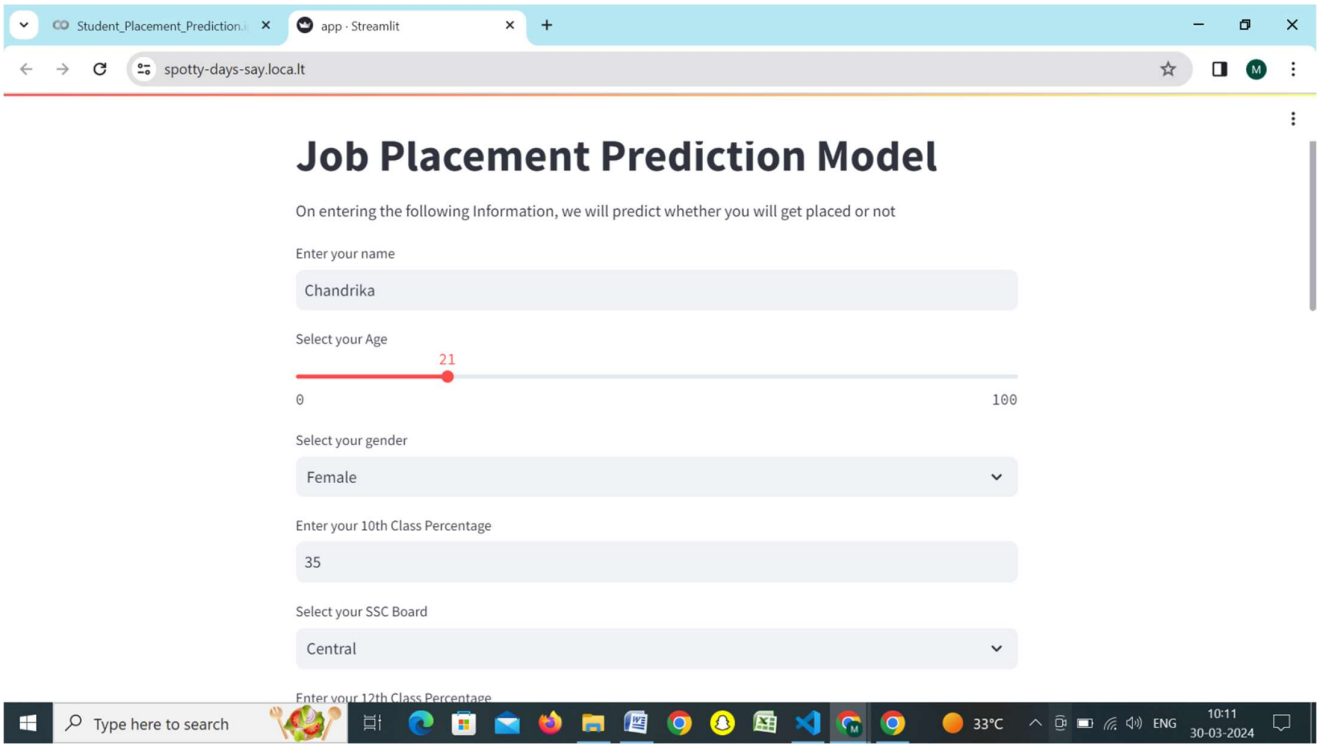
CHAPTER 11

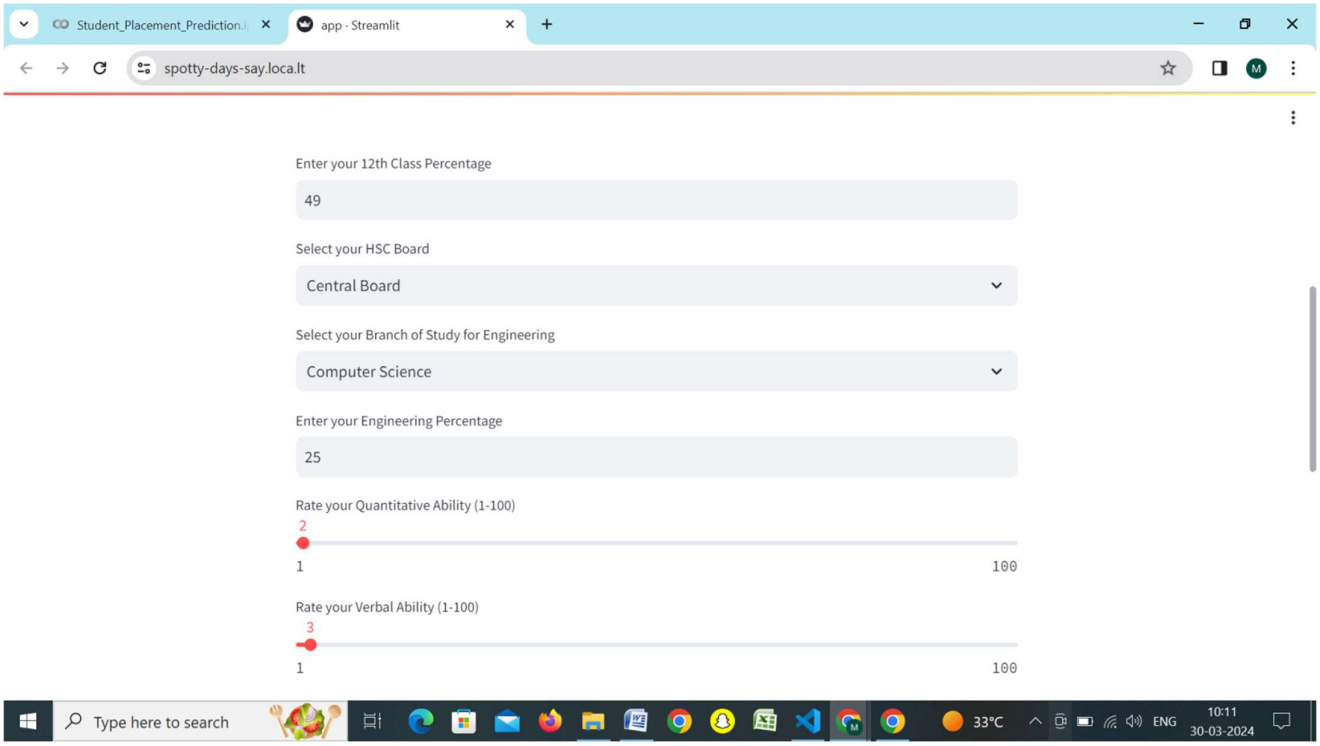
RESULTS

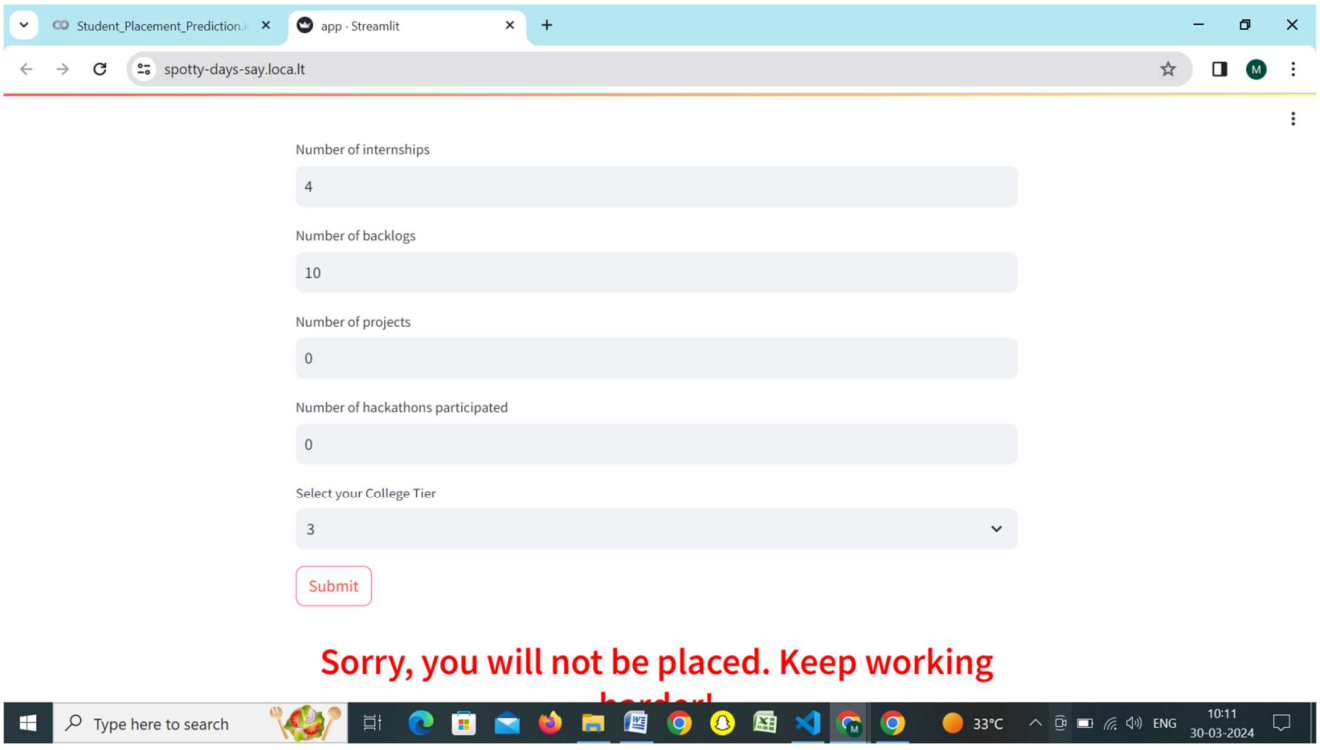


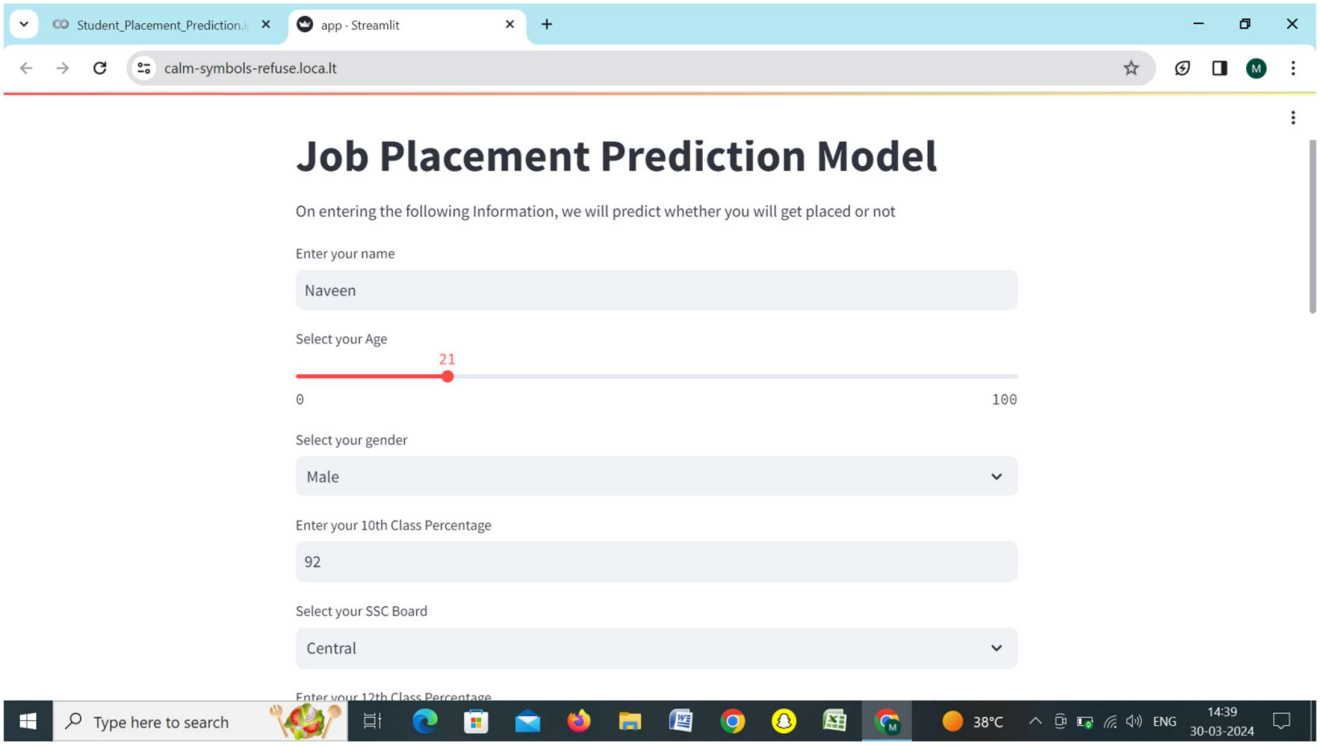


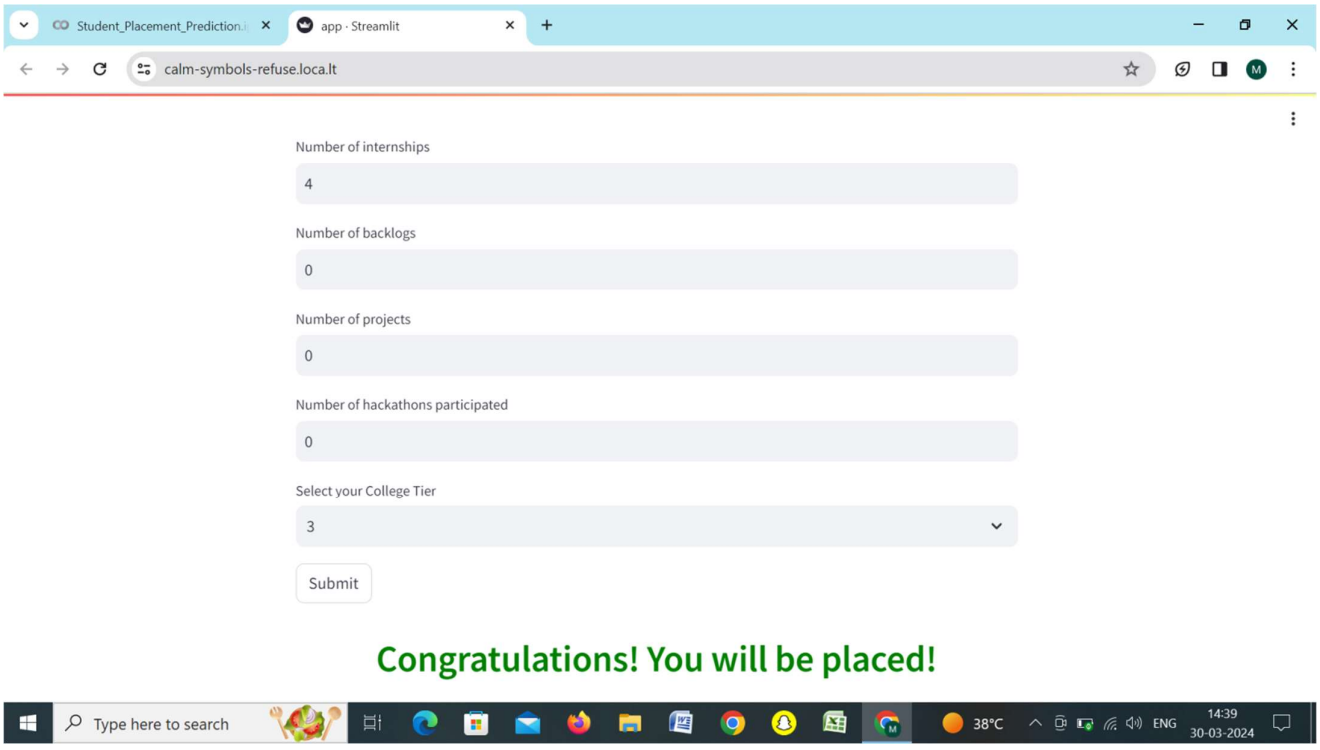

Student_Placement_Prediction
app · Streamlit
spotty-days-say.loca.lt
Rate your Verbal Ability (1-100)
25
1
100
Number of internships
4
Number of backlogs
0
Number of projects
0
Number of hackathons participated
0
Select your College Tier
3
Type here to search
33°C
ENG
10:06
30-03-2024

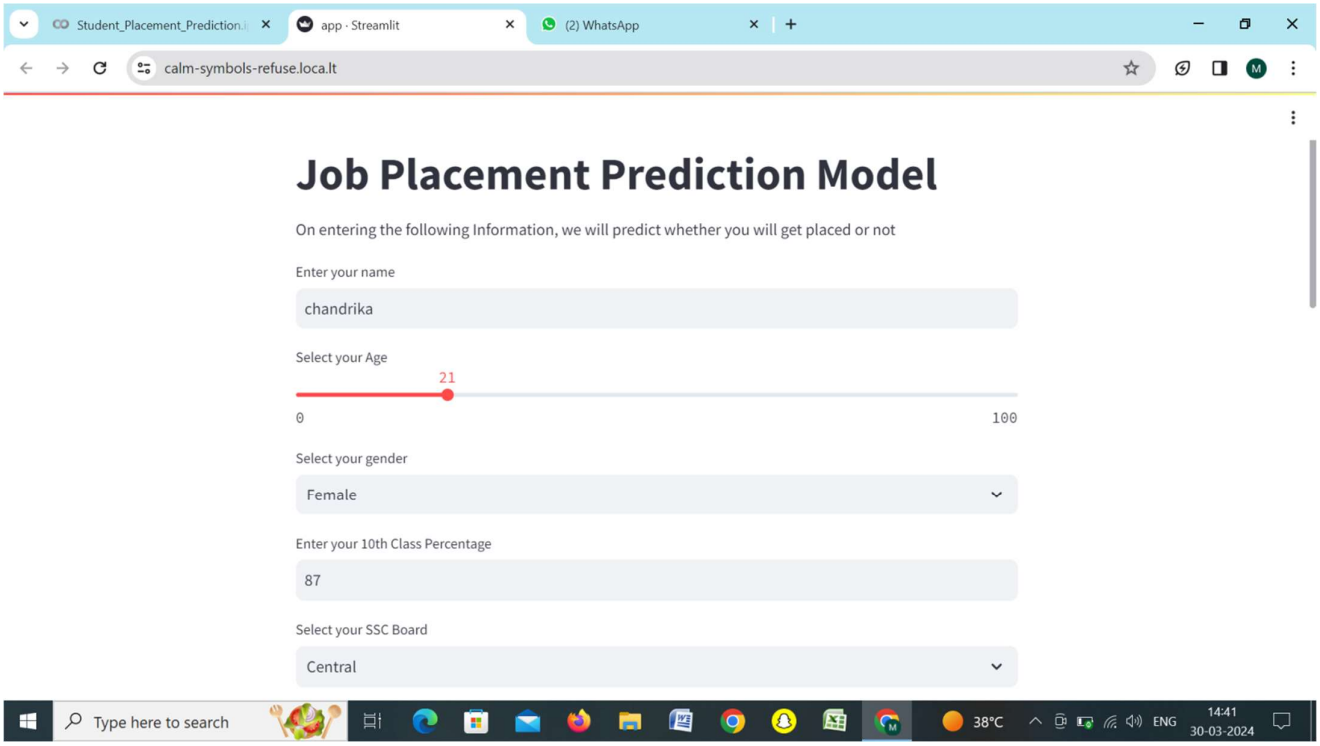

Student_Placement_Prediction
app · Streamlit
spotty-days-say.loca.lt
Job Placement Prediction Model
On entering the following Information, we will predict whether you will get placed or not
Enter your name
Chandrika
Select your Age
21
0
100
Select your gender
Female
Enter your 10th Class Percentage
35
Select your SSC Board
Central
Type here to search
33°C
ENG
10:11
30-03-2024

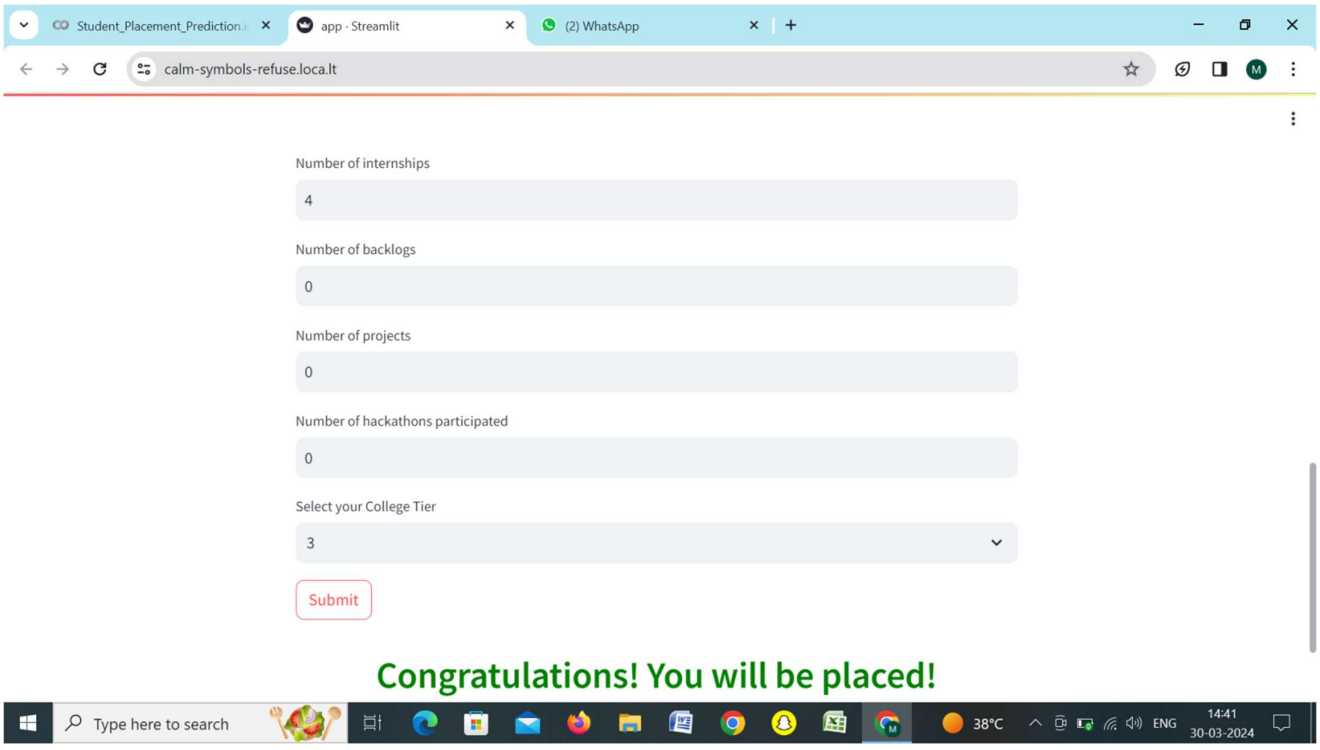

Student_Placement_Prediction
app · Streamlit
spotty-days-say.loca.lt
Enter your 12th Class Percentage
49
Select your HSC Board
Central Board
Select your Branch of Study for Engineering
Computer Science
Enter your Engineering Percentage
25
Rate your Quantitative Ability (1-100)
2
1
100
Rate your Verbal Ability (1-100)
3
1
100
Type here to search
33°C
ENG
10:11
30-03-2024


Student_Placement_Prediction
app · Streamlit
spotty-days-say.loca.lt
Number of internships
4
Number of backlogs
10
Number of projects
0
Number of hackathons participated
0
Select your College Tier
3
Submit
Sorry, you will not be placed. Keep working
Type here to search
33°C
ENG
10:11
30-03-2024


Student_Placement_Prediction
app · Streamlit
calm-symbols-refuse.loca.lt
Job Placement Prediction Model
On entering the following Information, we will predict whether you will get placed or not
Enter your name
Naveen
Select your Age
21
0
100
Select your gender
Male
Enter your 10th Class Percentage
92
Select your SSC Board
Central
Type here to search
38°C
ENG
14:39
30-03-2024


Student_Placement_Prediction
app · Streamlit
calm-symbols-refuse.loca.lt
Number of internships
4
Number of backlogs
0
Number of projects
0
Number of hackathons participated
0
Select your College Tier
3
Submit
Congratulations! You will be placed!
Type here to search
38°C
ENG
14:39
30-03-2024


Student_Placement_Prediction
app · Streamlit
(2) WhatsApp
calm-symbols-refuse.loca.lt
Job Placement Prediction Model
On entering the following Information, we will predict whether you will get placed or not
Enter your name
chandrika
Select your Age
21
0
100
Select your gender
Female
Enter your 10th Class Percentage
87
Select your SSC Board
Central
Type here to search
38°C
ENG
14:41
30-03-2024


Student_Placement_Prediction
app · Streamlit
(2) WhatsApp
calm-symbols-refuse.loca.lt
Number of internships
4
Number of backlogs
0
Number of projects
0
Number of hackathons participated
0
Select your College Tier
3
Submit
Congratulations! You will be placed!
Type here to search
38°C
ENG
14:41
30-03-2024

CHAPTER 12

CONCLUSION

The main motto of our project is to provide accurate results. Usually Logistic regression is a valuable tool for predicting student placement success. By analyzing historical data on past students' academic performance, extracurricular activities, or other relevant factors, the model can estimate the likelihood of future students landing a job. This information can be used by universities to guide students towards areas where they excel and by companies to target their recruitment efforts. It's important to remember that logistic regression is a statistical model and it is clear that our model predict the results accurately.

CHAPTER 13

REFERENCES

[1] “Prediction Model for Students Future Development by Deep Learning and TensorFlow Artificial Intelligence Engine” 2018 4th IEEE International Conference on Information Management

[2] Dr. A. Padmapriya (November 2012) Prediction Of Higher Education Admissibility Using Classi fication Algorithms International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X

[3] H. Sabnani, M. More, P. Kudale, S. Janrao, “Prediction of Student Enrolment Using Data Mining Techniques”, International Research Journal of Engineering and Technology (IRJET), 5(4), 1830-18 33, 2018.

[4] Animesh,G.,M,Vignesh.,Bysani,P.,Naini,D., 2015. A Placement Prediction System Using K-Near est Neighbors Classifier. 2nd International Conference on Cognitive Computing and Information Pro cessing (CCI).

[5] Karan. P.,Prateek, B., 2015. “Application of Data mining in predicting student placement”, Intern ational conference on Green computing and Internet of things (ICGCIoT).).

[6] Patel, T., Tamrakar, A. “Data Mining Techniques for Campus Placement Prediction in Higher Ed ucation” Indian J.Sci.Res. 14 (2) 2017: 467-471.).

[7] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar,PPS – Placement Prediction System using Logistic Regression in 2014 IEEE International Conference on MOOC, Innovation and Technology in Education.

[8] S.Taruna , Mrinal Pandey ,An Empirical Analysis of Classification Techniques for Predicting Academic Performance in 2014 IEEE International Advance Computing Conference (IACC).

[9] Kotsiantis, Sotiris B., and Panayiotis E. Pintelas, "Predicting students marks in hellenic open university", in Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on, pp. 664-668. IEEE, 2005.

[10] Syed Tanveer Jishan, Raisul Islam Rashu, Naheena Haque and Rashedur M Rahman, Improving accuracy of students final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, in Decision Analytics (2015) 2:1 DOI 10.1186/s40165-014- 0010-2(Springer Journel).

[11] Zhiwu Liu and Xiuzhi Zhang, Prediction and Analysis for Students Marks Based on Decision Tree Algorithm in 2010 Third International Conference on Intelligent Networks and Intelligent Systems.

[12] Hitarthi Bhatt, Shraddha Mehta, Lynette R. D'mello, Use of ID3 Decision Tree Algorithm for Placement Prediction, in (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4785-4789.