VISVESVARAYA TECHNOLOGICAL UNIVERSITY
JNANA SANGAMA, BELAGAVI – 590 018

An Internship Report on

"PREDICTING STUDENTS UNIVERSITY ADMISSION USING LOGISTIC REGRESSION"

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

Submitted by:
Dhanvantri Kulkarni
2KA20IS009

Under the Guidance of

Dr. Rajashekar Kunabeva
Associate Professor, Head
Dept. of ISE,
SKSVMACET, Lakshmeshwar.

Department of Information Science and Engineering

Smt. Kamala & Sri. Venkappa M. Agadi College of Engineering & Technology

Lakshmeshwar-582 116

2023-2024

Certificate

This is to certify that the internship work entitled "Predicting Students University Admission Using Logistic Regression" is bonafied work carried out by Dhanvantri Kulkarni(2KA20IS009), in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering in Information Science and Engineering of Visvesvaraya Technological University, Belagavi**, during the year 2023-2024. It is certified that all the corrections/suggestions indicated for internal assessment have been incorporated in the report. The internship report has been approved as it satisfies the academic requirements in respect of internship work prescribed for the Bachelor of Engineering degree.

Signature of the Guide

Dr.Rajashekar Kunabeva
Associate Professor,Head
Dept.of ISE
SKSVMACET
Lakshmeshwar

Signature of HOD

Dr. Rajashekar Kunabeva
Asst.Professor and Head ,
Dept.of ISE,
SKSVMACET
Lakshmeshwar

Signature of the principal

Dr. Parashuram Baraki
Principal
SKSVMACET
Lakshmeshwar

Examiners:
1.
2.

I

DECLARATION

I, **Dhanvantri Kulkarni** bearing the USN **2KA20IS009** studying in the seventh semester of Bachelor of Engineering in Information Science and Engineering at Smt. Kamala & Sri.Venkappa M. Agadi College of Engineering & Technology ,Lakshmeshwar, hereby declare that this internship work entitled "**Predicting Student University Admission Using Logistic Regression**" which is being submitted by me in the partial fulfillment for the award of the degree of Bachelor of Engineering in Information Science and Engineering, from Visvesvaraya Technological University, Belagavi is an authentic record of me carried out during the academic year 2023-2024, under the guidance of **Dr. Rajashekar Kunabeva** Asst. Prof,Head of Department of Information Science & Engineering, Smt. Kamala & Sri. Venkappa M. Agadi College of Engineering & Technology, Lakshmeshwar.

I further undertake that the matter embodied in the dissertation has not been submitted previously for the award of any degree by me to any other university or institution.

Place:Lakshmeshwar   **Dhanvantri Kulkarni**

Date:

ACKNOWLEDGEMENT

It is my proud privilege and duty to acknowledge the kind of help and guidance received from several people in preparation of this internship report. It would not have been possible to prepare this report in this form without their valuable help, cooperation and guidance.

I wish to record my sincere gratitude to my internship guide **Dr. Rajashekar Kunabeva,**Information Science and Engineering Department for guiding me in investigations for this internship work and providing encouragement, constant support and guidance which was of a great help to complete this seminar successfully.

I am grateful to **Dr. Rajashekar Kunabeva**, Head of the Department of Information Science and Engineering for giving me the support and encouragement that was necessary for the completion of this internship work.

I would also like to express my gratitude to **Dr. Parashuram Baraki,** Principal, for providing us pleasant environment to work in library and for providing laboratory facilities needed to prepare this report.

Last but not the least, we wish to thank our **parents** for financing our studies in this college as well as for constantly encouraging us to learn engineering. Their personal sacrifice in providing this opportunity to learn engineering is gratefully acknowledged.

Place: Lakshmeshwar   **Dhanvantri Kulkarni**

ABSTRACT

The primary purpose is to discuss the prediction of student admission to university based on numerous factors and using logistic regression. Many prospective students apply for Master's programs. The admission decision depends on criteria within the particular college or degree program. The independent variables in this study will be measured statistically to predict graduate school admission. Exploration and data analysis, if successful, would allow predictive models to allow better prioritization of the applicants screening process to Master's degree programme which in turn provides the admission to the right candidates.
.

TABLE OF CONTENTS

CHAPTER 1

COMPANY PROFILE

1.1 INTRODUCTION

Fig 1.1: Company Logo

AiRobosoft is a Software, Robotics combined EV Manufacturing Company with HQ and development centre in Bangalore, India. Their Product's   Include development's using technologies such as Artificial Intelligence, Cloud computing Machine learning, Embedded Systems and Internet of Things.

AiRobosoft was founded on the robust idea:We all live in a society where we keep adapting things or culture.with the last technological trends,our focus is to make products advancements to adaptability for serving society.

1.2 CONTACT DETAILS

Mail : support@airobosoft.com

Tel no : +91-88843 99 089

          +91-99644 83 843.

Linked In : AiRobosoft

Web : www.airobosoft.com

Name of founder: Syed Asad Ahmed, CEO and Founder, AiRobosoft.

1.3 COMPANY LOCATION

AiRobosoft Products and Services

No - 4, 3rd Floor, 5th A Main Rd, Adjacent to Bangalore Baptist Hospital, Vinayakanagar, Hebbal, Bengaluru, Karnataka - 560024.

1.4 INTERNSHIP DETAILS

Internship Scheduled Period: 1 Month

Start of Internship:14/08/2023

End of Internship :09/09/2023

Mode of Internship: Offline

CHAPTER 2

ABOUT COMPANY

2.1 VISION

As the pollution increases and natural resources depleting, with the increase in fuel rates.

Our Vision is to become a trusted EV Brand with a Global Presence, making INDIA as a complete Electric Vehicle enabled Country.

Core Values:-

Dependability: We aim to keep all our products performance to be able to accomplish its assigned mission.

Consistency: We work to keep all the technologies consistently in reach for customer use and satisfaction.

Ownership: We are accountable for all the actions we take and keep customer feelings in due regard.

Integrity: We provide good value and satisfy customers expectation, being ethical, fair and transparent.

2.2 COOPERATION MODEL

2.2.1 Fixed Price model

- Well defined scope.
- Finalized pricing
- Sticking to Budget.
- Fast Delivery
- Predictability

2.2.2 Time and Material model

- Vague Scope.
- Flexibility.
- Constant Improvements.

2.2.3 Achievement model

- Controlled Results
- Paying on Achievements
- Constant Improvements
- Predictability

2.3 PRODUCT AND SERVICES

2.3.1 Automotive

comprises a wide range of companies and organizations involved in the design, development, manufacturing, marketing, selling, repairing, and modification of motor vehicles.[1]

2.3.2   Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

2.3.3 Data Science

Data science deals with data extraction and generating meaningful insights from it using statistics, machine learning and other computational tools.

2.3.4   Embedded System

Embedded System is an integrated system that is formed as a combination of computer hardware and software for a specific function.

2.3.5   IoT

The Internet of things (IoT) describes devices with sensors, processing ability, software and other technologies that connect and exchange data with other devices and systems over the Internet or other communications network.

2.3.6 E Learning

E-learning, also referred to as online learning or electronic learning, is the acquisition of knowledge that takes place through electronic technologies and media.

2.3.7 Industrial Automation

Industrial Automation is a process of operating machines and other industrial equipment with the help of digital logical programming and reducing human intervention in decision making and manual command process with the help of mechanized equipment.

2.3.8 Data Analytics

Data Analytics is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making.

CHAPTER 3

TASKED PERFORMED

3.1 INTRODUCTION TO PYTHON

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Python is a great general-purpose programming language on its own, but with the help of a few popular libraries (numpy, scipy, Matplotlib) it becomes a powerful environment for scientific computing.

Python Version

Python 2

Published in late 2000, Python 2 signalled a more transparent and inclusive language development process than earlier versions of Python with the implementation of PEP (Python Enhancement Proposal), a technical specification that either provides information to Python community members or describes a new feature of the language.

Additionally, Python 2 included many more programmatic features including a cycle-detecting garbage collector to automate memory management, increased Unicode support to standardize characters, and list comprehensions to create a list based on existing lists. As Python 2 continued to develop, more features were added, including unifying Python's types and classes into one hierarchy in Python version 2.2.

Python 3

Python 3 is regarded as the future of Python and is the version of the language that is currently in development. A major overhaul, Python 3 was released in late 2008 to address and amend intrinsic design flaws of previous versions of the language. The focus of Python 3 development was to clean up the code base and remove redundancy, making it clear that there was only one way to perform a given task.

Major modifications to Python 3.0 included changing the print statement into a built-in function, improve the way integers are divided, and providing more Unicode support.

Python 2.7

Following the 2008 release of Python 3.0, Python 2.7 was published on July 3, 2010 and planned as the last of the 2.x releases. The intention behind Python 2.7 was to make it easier for Python 2.x users to port features over to Python 3 by providing some measure of compatibility between the two. This compatibility support included enhanced modules for version 2.7 like unit test to support test automation, argparse for parsing command-line options, and more convenient classes in collections.

Key Differences

While Python 2.7 and Python 3 share many similar capabilities, they should not be thought of as entirely interchangeable. Though you can write good code and useful programs in either version, it is worth understanding that there will be some considerable differences in code syntax and handling.
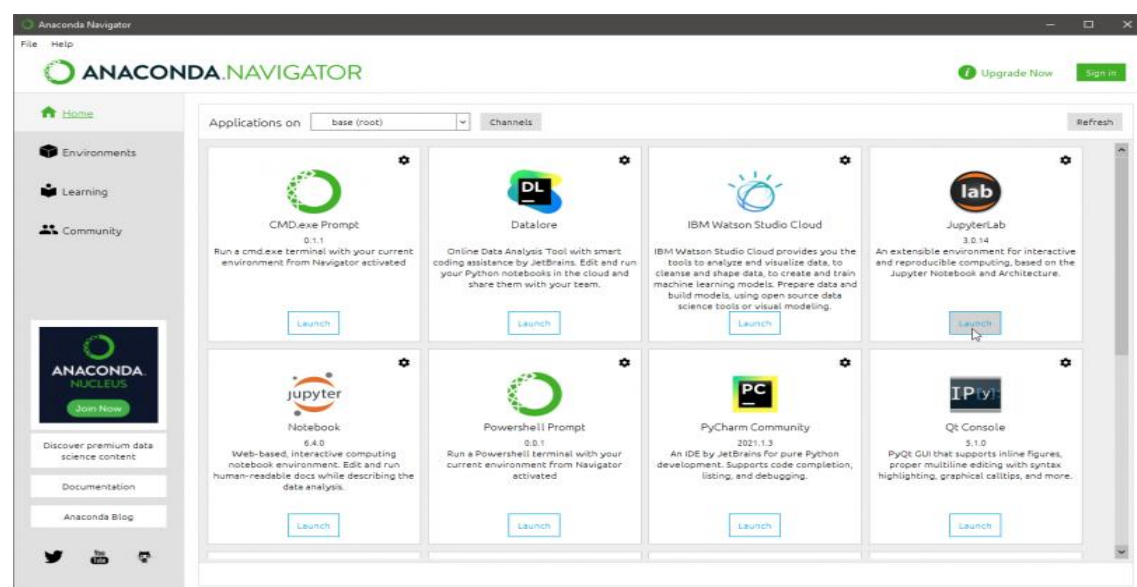
3.2 INTRODUCTION TO ANACONDA IDE

Anaconda is an open-source package manager, environment manager, and distribution of the Python and R programming languages. It is commonly used for large-scale data processing, scientific computing, and predictive analytic, serving data scientists, developers, business analysts, and those working in DevOps.

Anaconda offers a collection of over 720 open-source packages, and is available in both free and paid versions. The Anaconda distribution ships with the conda command-line utility.

Why Anaconda?

- User level install of the version of python you want
- Able to install/update packages completely independent of system libraries or admin privileges
- conda tool installs binary packages, rather than requiring compile resources like pip - again, handy if you have limited privileges for installing necessary libraries.
- Comes either in full-meal-deal version, with numpy, scipy, PyQt, spyder IDE, etc. or in minimal / alacarte version (mini conda) where you can install what you want, when you need it.

.Fig 3.1: Snapshot of Anaconda IDE

3.3 INTRODUCTION TO JUPYTER NOTEBOOK

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

● **Installation**

use a handy tool that comes with Python called **pip** to install Jupyter Notebook

 $ pip _install_ jupyter

● **Starting jupyter notebook server**

To get started, all you need to do is open up your terminal application and go to a folder of your choice. I recommend using something like your Documents folder to start out with and create a sub folder there called *Notebooks* or something else that is easy to remember.
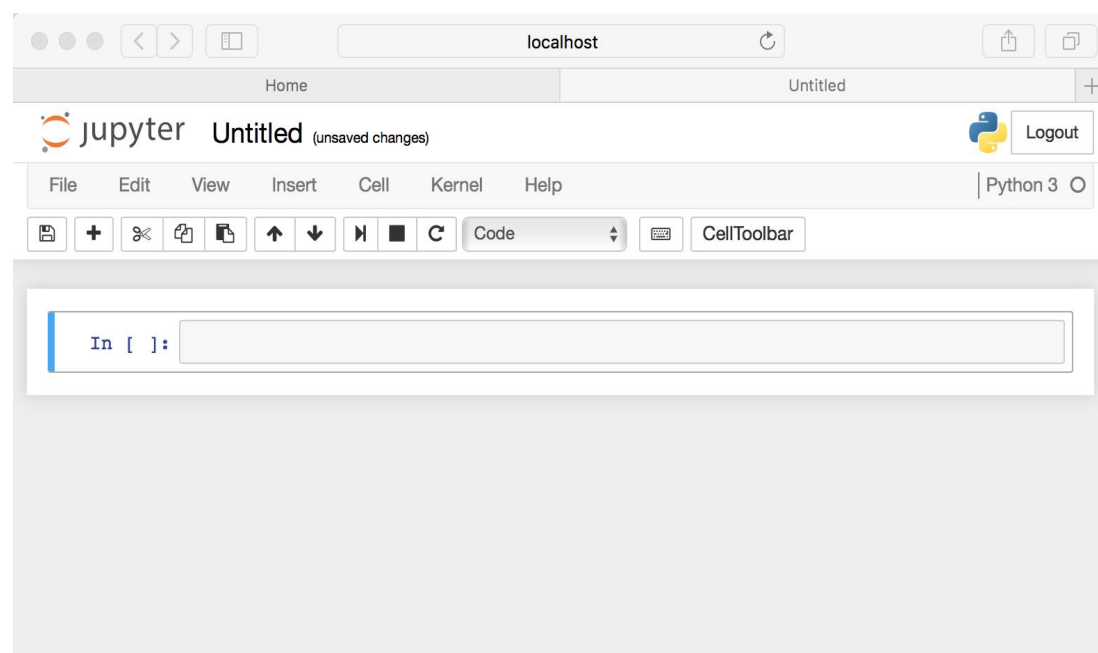
$ jupyter notebook

Fig 3.2: Snapshot of Jupyter Notebook

● **Running Cells**

A Notebook's cell defaults to using code whenever you first create one, and that cell uses the kernel that you chose when you started your Notebook.

In this case, you started yours with Python 3 as your kernel, so that means you can write Python code in your code cells. Since your initial Notebook has only one empty cell in it, the Notebook can't really do anything.
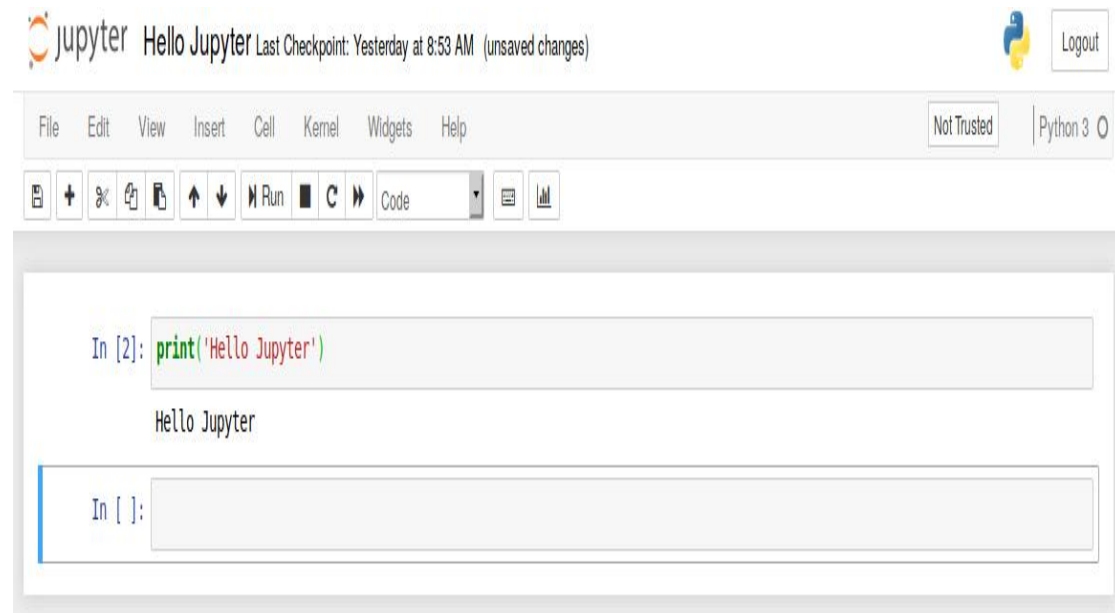

Fig 3.3: Snapshot of running cell

3.1.1 Modules Used

● **NumPy**

The NumPy package provides access to a new data structure called arrays which allow efficient vector and matrix operations. It also provides a number of linear algebra operations (such as solving of systems of linear equations, computation of Eigenvectors and Eigenvalues).

Fig 3.4: Numpy Module

● **Pandas**

Pandas is a Python package providing fast, flexible, and expressive data structures designed to work with *relational* or *labeled* data both. It is a fundamental high-level building block for executing practical, real world data analysis in Python.

```
In [6]:  import pandas as pd
         pd.DataFrame(data = [
         ['Steven', 'Towaco', 20],
         ['Janet', 'San Francisco', 25],
         ['Laura', 'Austin', 38],
         ['Anne', 'Baltimore', 12],
         ],
                      columns = ['Name', 'City', 'Age'])
         pd.Series(data=['Steven', 'Janet', 'Laura', 'Anne'])

Out[6]:  0      Steven
         1       Janet
         2       Laura
         3        Anne
         dtype: object
```

Fig 3.5: Pandas Module

● **Matplotlib**

Matplotlib is the most popular and mature library for plotting data using Python. It has all of the functionality you would expect, including the ability to control the formatting of plots and figures at a very fine level.
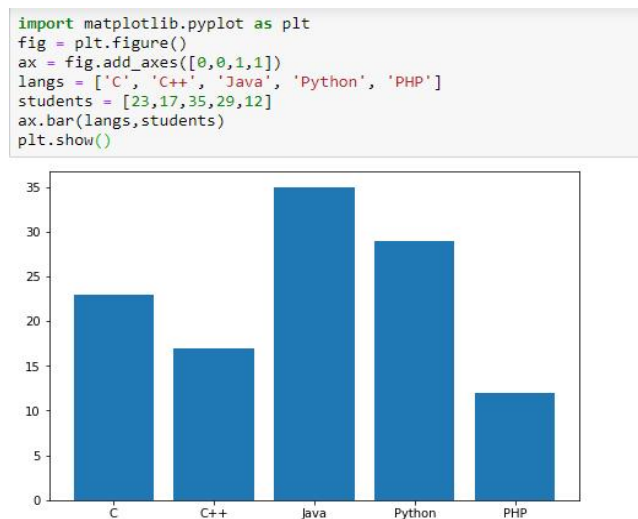
```
import matplotlib.pyplot as plt
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
langs = ['C', 'C++', 'Java', 'Python', 'PHP']
students = [23,17,35,29,12]
ax.bar(langs,students)
plt.show()
```


Fig 3.6: Matplotlib Module

3.4 LOGISTIC REGRESSION

Logistic regression is used when there is a binary 0-1 response, and potentially multiple categorical and/or continuous predictor variables. Logistic regression can be used to model probabilities (the probability that the response variable equals 1) or for classification.

Sigmoid Function

Now we use the <u>sigmoid function</u> where the input will be z and we find the probability between 0 and 1. i.e predicted y.
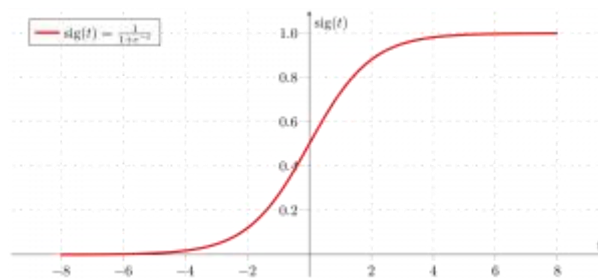
$$\sigma(z) = \frac{1}{1-e^{-z}}$$

Fig 3.7: Sigmoid Curve

Logistic Regression Equation

$$\frac{p(x)}{1-p(x)} = e^z$$

Applying natural log on odd. then log odd will be

$$\log \left[\frac{p(x)}{1-p(x)}\right] = z$$
$$\log \left[\frac{p(x)}{1-p(x)}\right] = w \cdot X + b$$

then the final logistic regression equation will be:

$$p(X;b,w) = \frac{e^{w \cdot X+b}}{1+e^{w \cdot X+b}} = \frac{1}{1+e^{-w \cdot X+b}}$$

Types of Logistic Regression

Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"

Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

3.5 SYSTEM DESIGN

3.5.1 Loading Dataset

You can get the .csv file as a matrix in python this function can also read .txt files too. However, we need to use the pandas package and it may increase the complexity usually.

The dataset contains several columns which we can use as predictor variables:

- gpa
- gre score
- rank or prestige of an applicant's undergraduate alma mater
- The fourth column, admit, is our binary target variable. It indicates whether or not a candidate was admitted our not.

```
import pandas as pd
import statsmodels.api as sm
import pylab as pl
import numpy as np
# read the data in
df = pd.read_csv("binary.csv")
# df = pd.read_csv("https://stats.idre.ucla.edu/stat/data/binary.csv")
df.head
```
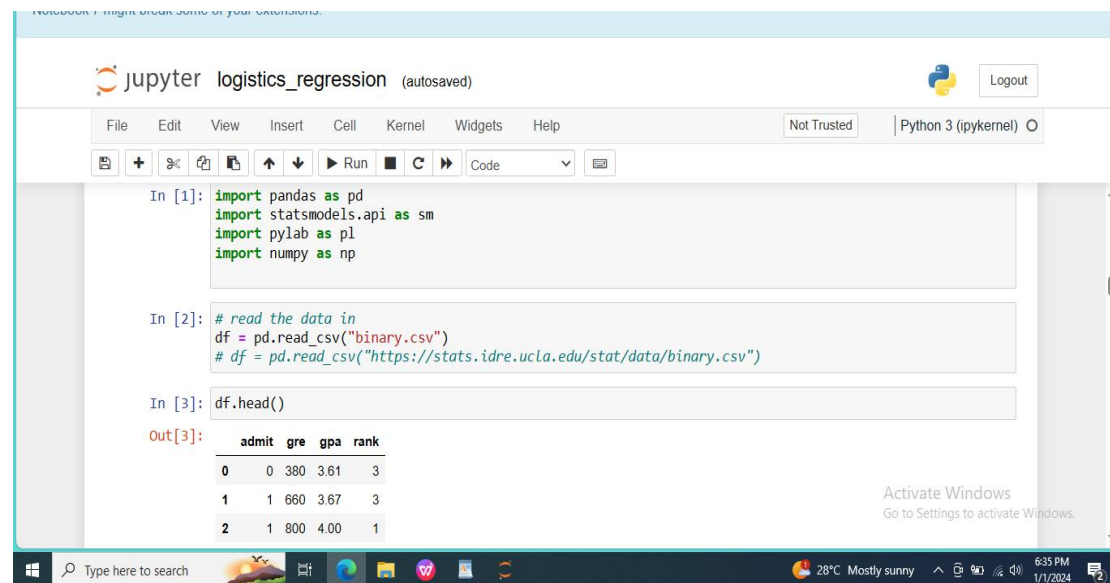
Fig 3.8: Loading Dataset

3.5.2 Summarising the Dataset

This is find how will be the count of admission when the GRE,GPA and prestige is max and min and this can be done by using the below code by running this code we will get the percentage of admission that for which variable it is maximum and for which it is minimum.
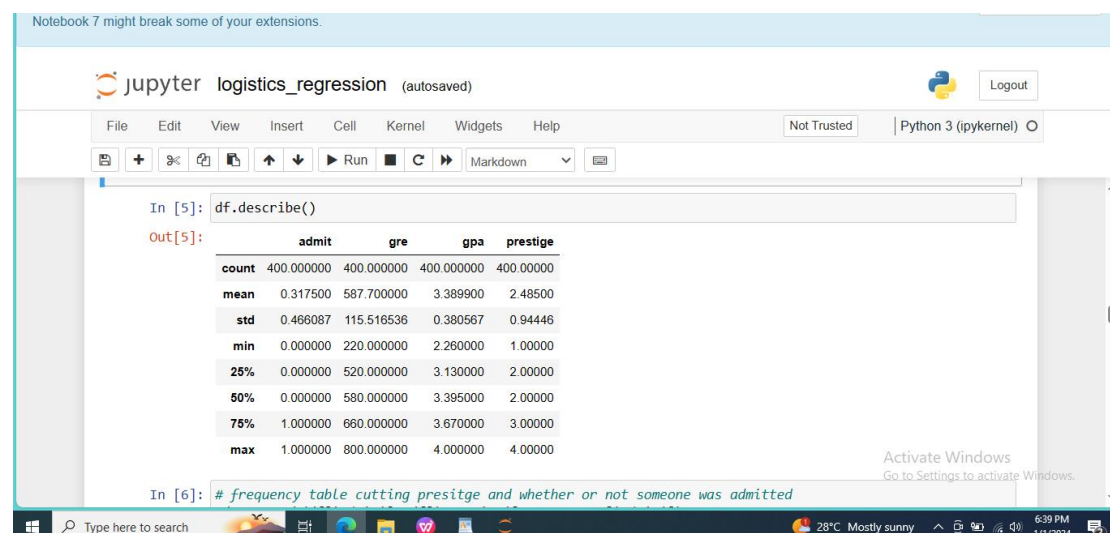
df.describe()

Fig 3.9: Summarising the Dataset

3.5.3 Display the variable

Here we will be displaying the prestige variable for cross checking that the admission are based on prestige variable and to display the prestiage variable we will running the below code

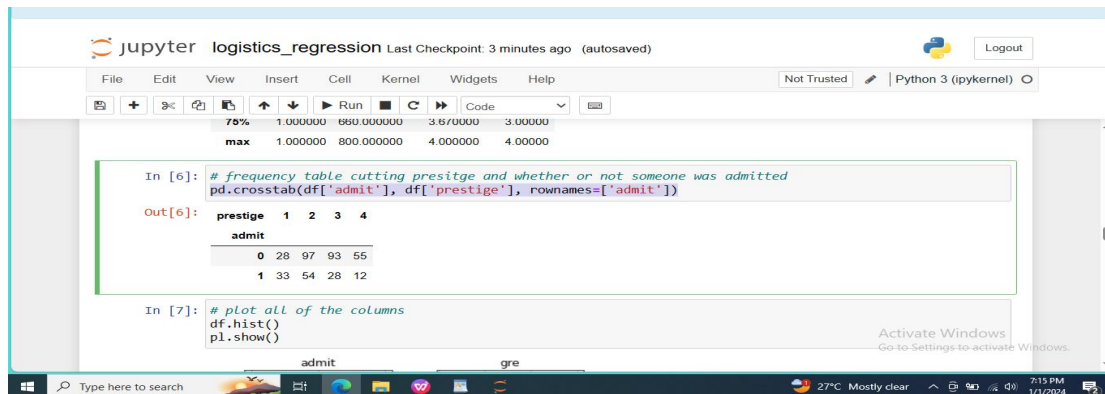pd.crosstab(df['admit'], df['prestige'], rownames=['admit'])

Fig 3.10: Displaying the Variable

3.5.4 Result

Here we will be plotting all the variable using the below code.From the plots it has clearly observed that chance of admission is high when somebody belongs to high ranking university. Although some students are from average rating university,still they have a chance to get admitted.
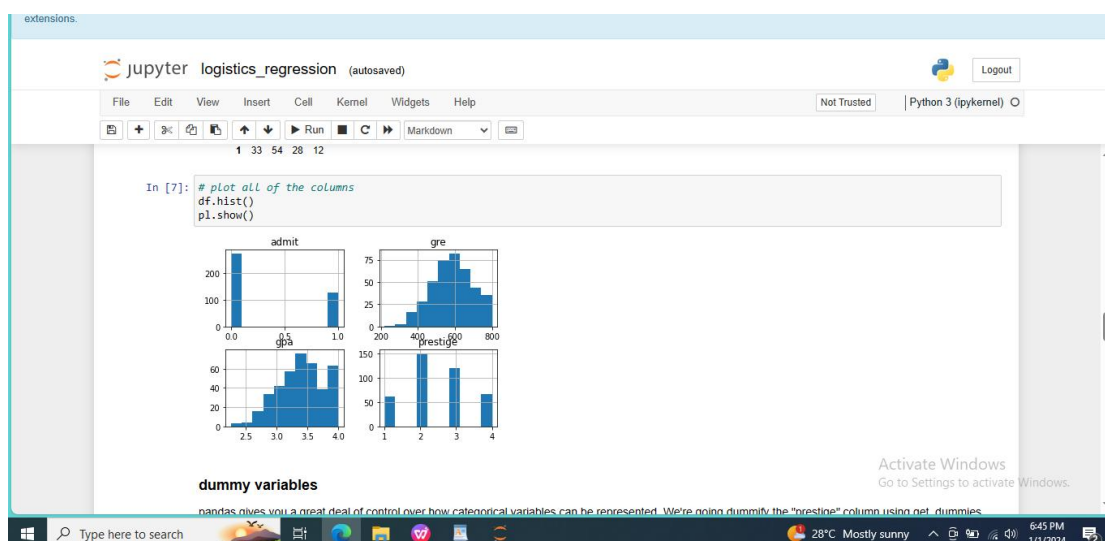

Fig 3.11 :Plot of All Variables

CHAPTER 4

REFLECTION

4.1 EFFECTIVE COMMUNICATION

Effective communication is the process of exchanging ideas, thoughts, opinions, knowledge, and data so that the message is received and understood with clarity .

4.2 NEGOTIATION

Negotiation is a dialogue where two or more sides work together to reach an agreeable solution for all involved. It might result in a formal agreement, like a contract, or a less formal understanding, like a verbal agreement.

4.3 CRITICAL THINKING

Critical thinking is the ability to effectively analyze information and form a judgment.To think critically, you must be aware of your own biases and assumptions.

4.4 DECISION MAKING

Decision-making is an important skill that all employees need. Proper decision making techniques can help guarantee you handle problems in the best manner.

4.5 TIME MANAGEMENT

Time management is the process of consciously planning and controlling time spent on specific tasks to increase how efficient you are.

4.6 STRATEGIC PLANNING

Strategic planning process requires considerable thought and planning on the part of a company's upper-level management.

CHAPTER 5

CONCLUSION

The results of this examination appear to indicate that it greatly contributes to the response variable 'Chance of Admit'. Higher the GRE,Prestige score then higher the admit chances. The model predicts 87.5% accuracy and can be used for predicting the admit chances based on the above factors. This model will be helpful for the universities to predict the admission and ease their process of selection and timelines.As part of the hypothesis, the model proved that admission to Master's degree program is dependent on GRE,prestige and other scores. This model would likely be greatly improved by the gathering of additional data of students from different universities which has similar selection criteria to choose the candidates for Master's program.

REFERENCE

[1]   Mohan S Acharya, Dec 30, 2018, "A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019".

[2]   Raw data retrieve   from
      https://www..kaggle.com/mohansacharya/graduate-admission

[3]   project - 2 - google drive

[4]   Python Day 1 - Jupyter Notebook

[5]   logistics_regression - Jupyter Notebook

[6]   Numpy Tutorial - Jupyter Notebook

[7]   Pandas+Tutorial - Jupyter Notebook

[8]   Matplotlib - Jupyter Notebook

[9]   www.google.com

[10]   Internship Projects - Google Drive

CERTIFICATE OF COMPLITION

Reg. No: AAS7147

AiRobosoft
Our Vision, Your Future.

Internship Certification

This is to certify that

Dhanvantri Kulkarni

Has completed internship at **AIROBOSOFT PRODUCTS AND SERVICES LLP**
On
Artificial Intelligence and Machine Learning

August 14th 2023 – September 14th 2023

Her / His performance was Very Good.

(Scan to Authenticate)

Managing Director