IIIT ADMISSION PREDICTION

TEAM:

G.GANGADHAR (R170061)

D.JAYAKRISHNA (R170064)

TABLE OF CONTENTS

ABSTRACT

- IIIT admission Prediction is very important and plays a key role in Andhra Pradesh students who just passed 10th class. Predicting IIIT admission can be especially difficult because the students are not aware of admission requirements. For that reason, the main purpose of this research work is to provide a recommender system for early predicting IIIT admission.
- Therefore, To predict chance of admission first we applied several Supervised Machine Learning algorithms namely Linear Regression, Support Vector Regression, Decision Tree Regression, and Random Forest Regression. Secondly we compared and evaluated algorithms used to create a predictive model based on various evaluation metrics. Lastly we determined the most important parameters that influence the chance of admission.
- The experimental results showed that the Linear Regression is the most suitable Machine Learning algorithm for predicting IIIT admission

INTRODUCTION

- The number of students applying for IIIT(RGUKT) has increasing year by year. This fact has motivated us to study the grades of students and the chance of admission for IIIT that can help students in predicting the possibility of getting admission in IIIT.

- Machine Learning is a subset of Artificial Intelligence (AI) that enables computers to automatically improve through experience. In the area of education, the adoption of Machine Learning is also accelerating.

- Using the machine learning algorithms it's easy to build the prediction model based on the previous year's admission data.

LITERATURE SURVEY

REGRESSION

- **Regression** is a technique of **Supervised Algorithms** for investigating the relationship between independent and a dependent variable. It's used as a method for predictive modelling in Machine Learning, in which an algorithm is used to predict continuous outcomes.
- As we are predicting the percentage of chance of admission , the output(Dependent) variable is continuous variable. We have to use the **Regression** techniques to predict the continuous variable

Types of Regression Algorithms used for Prediction:

LINEAR REGRESSION(LR) : It is the most important algorithm in the field of Machine Learning, especially supervised learning. It is a way to model a relationship between a dependent variable and one or more independent variables. It consists of finding a regression line straight line through the points.

LITERATURE SURVEY

RANDOM FOREST REGRESSION(RFR) : it is an ensemble learning method that constructs a multitude of decision trees at training time and uses the average prediction of the individual trees to improve the prediction

SUPPORT VECTOR REGRESSION (SVR) : It is also a very popular Machine Learning technique used in both classification and regression. It is similar to Linear Regression with only a few minor differences. SVR allows defining how much error is acceptable in our predictive model and will find an appropriate line to fit the data

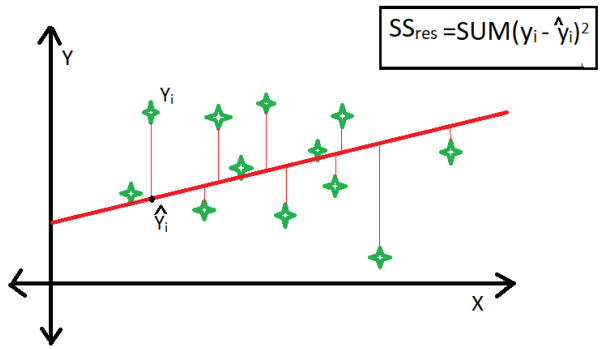
DECISION TREE REGRESSION (DTR) : It is the most widely used classification and prediction technique. It is a tree structure, where each internal node with outgoing edges indicates a condition on an attribute, each branch is an outcome of the test, and each leaf terminal node represents a class label

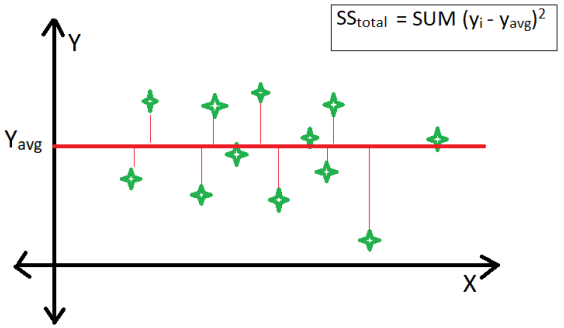
EVALUTION METHODS

The main part of building a Machine Learning model is Evaluation of model. There are many methods of evaluation that can be used. While building our model we have used the below two metrics for the evaluation part.

R-Squared (R^2 or the coefficient of determination) : is an indicator that allows judging the quality of simple linear regression. It measures the fit between the model and the observed data or how well the regression equation is to describe the distribution of points. The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.

$$R^2 = \frac{SS_{res}}{SS_{tot}}$$

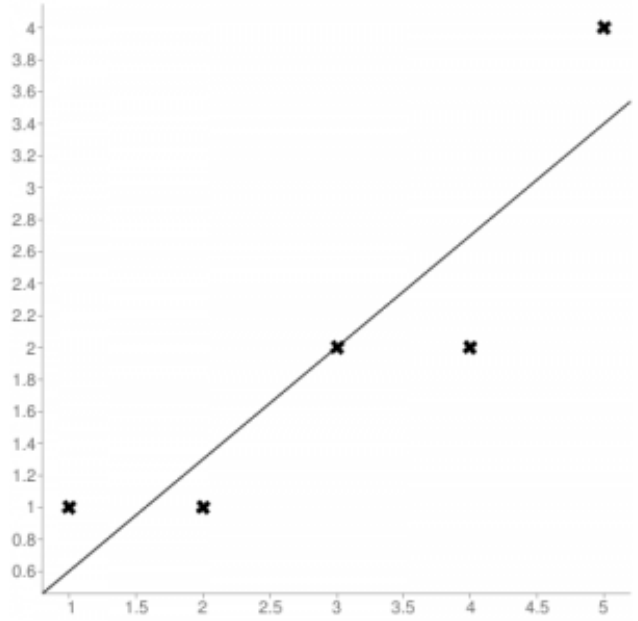




EVALUTION METHODS

Mean Square Error (MSE) : is the arithmetic mean of the squares of the predictions between the model and the observations. This is the value to be minimized in the context of a single or multiple regressions. It measures the average of error squares i.e. the average squared difference between the estimated values and true values. It is a risk function, corresponding to the expected value of the squared error loss.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$



INITIAL RESULTS OF ALGORITHMS

ALGORITHM	R2 (R SQUARED)	MEAN SQAURE ERROR
Support Vector Regression	0.8513	0.0792
Linear Regression	**0.9417**	0.0559
Random Forest Regression	0.9125	0.0492
Decision Tree Regression	0.8913	0.0743

From table, we can see that the **Linear Regression** suits best for our model.

LINEAR REGRESSION

- Linear regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used

- Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

TYPES OF LINEAR REGRESSION

Linear regression can be further divided into two types of the algorithm:

- **Simple Linear Regression**

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- **Multiple Linear regression:**

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

HYPOTHESIS FUNCTION

For the Linear Regression, the hypothesis function is defined by

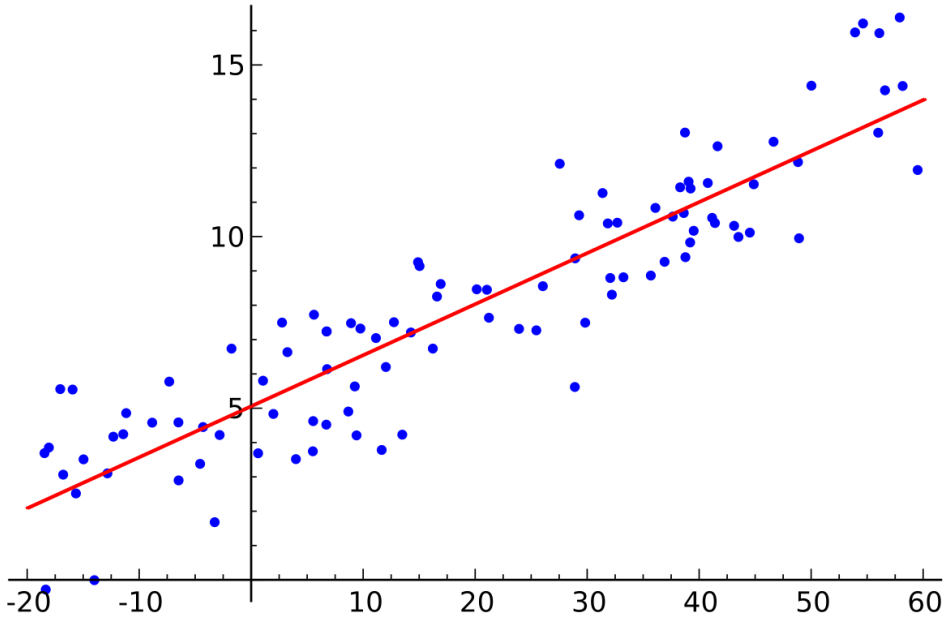
$$y = \theta_1 + \theta_2.x$$

Where,

X: input training data
y: output label
θ_1: intercept
θ_2: coefficient of x

LINEAR REGRESSION GRAPH

- When we plot the graph between dependent and independent variable , the Linear Regression finds a linear relationship between both variables and draws a perfect linear line.



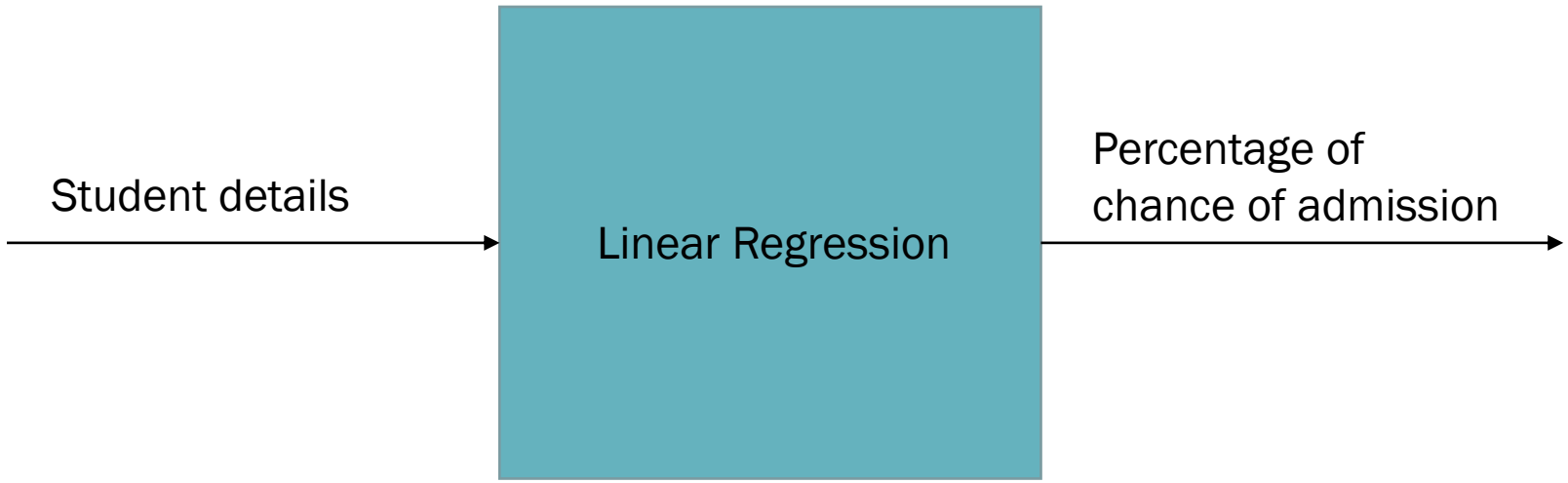
PROPOSED MODEL

- The students who completed tenth class have more curiosity and confusion whether they get admission in IIIT or not. At present there is no system available to help these students.

- So in this model we have proposed ‘IIIT Admission Prediction System’ which predicts the percentage of chance of getting admission into IIIT.

- The recommender system is built in the following steps
  - Data Collection
  - Data Pre-processing
  - Model Building

WORKING MODEL



DATA COLLECTION

- Initially we have collected the previous data related to admissions of IIIT and stored it in CSV file.
- The dataset has following attributes :

ATTRIBUTE	TYPE
HALLTICKET	QUANTITATIVE
CGPA	QUANTITATIVE
GENDER	QUALITATIVE
CASTE	QUALITATIVE
SCHOOL TYPE	QUALITATIVE
MANDAL	QUALITATIVE
DISTRICT	QUALITATIVE

DATA PRE-PROCESSING

- It represents one of the most crucial steps in all Machine Learning projects because it involves formatting data, Improving data quality, feature engineering, and labelling
- Firstly we need to process the data such that there will be no null values and duplicate values.
- In the given data set there are so many parameters that are of type String. We need to convert this type into numeric values for further processing.
- After doing this, add a new parameter called '**Performance Score**' by adding **caste, gpa, school type and gender.**

```
#Calculating performance score with major factors which affects chance of admit

df["Performance Score"] = df["Govt/Private School"] + df["Caste"] + df["Gender"] + df["10th GPA"]
```

```
df.head()
```

	Hall ticket (10th class)	10th GPA	Govt/Private School	Caste	Gender	Mandal	District	State	Chance of Admit	Performance Score
0	1720104115	10.0	0.0	0.5	0.34	Kallur	Kurnool	Ap	0.83	10.84
1	1722136839	10.0	0.4	0.5	0.34	Galiveedu	Kadapa	Ap	0.98	11.24
2	1718124014	9.7	0.4	0.5	0.66	Madanapalli	Chittoor	Ap	0.93	11.26
3	1722136791	9.8	0.4	0.7	0.34	Galiveedu	Kadapa	Ap	0.95	11.24
4	1722107642	9.8	0.4	0.1	0.66	Vallur	Kadapa	Ap	0.90	10.96

MODEL BUILDING

- In this section we have used Linear Regression algorithm function which was predefined from sklearn library in python3 language
- We divided the dataset into two parts such as Test data and Train Data
- Now we trained the model using train data by fitting into Linear Regression Model.
- After fitting the train data into model we get the outcomes for test data.
- Now the model built successfully.

IMPLEMENTATION

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [20]:
```
d=pd.read_csv("C:\\Users\\dasar\\Desktop\\project\\admission.csv")
```

In [21]:
```
df=pd.DataFrame(d)
```

In [22]:
```
df.head()
```

Out[22]:

	Hall ticket (10th class)	10th GPA	Govt/Private School	Caste	Gender	Mandal	District	State	Chance of Admit
0	1720104115	10.0	Private	BC-B	Male	Kallur	Kurnool	Ap	0.83
1	1722136839	10.0	Govt school	BC-B	Male	Galiveedu	Kadapa	Ap	0.98
2	1718124014	9.7	Govt school	BC-B	Female	Madanapalli	Chittoor	Ap	0.93
3	1722136791	9.8	Govt school	SC	Male	Galiveedu	Kadapa	Ap	0.95
4	1722107642	9.8	Govt school	OC	Female	Vallur	Kadapa	Ap	0.90

In [23]:
```
# Replacing Government schools with 0.4 and Private schools with 0 values

df['Govt/Private School']=df['Govt/Private School'].replace(['Govt school','Private'],[0.4,0])
```

In [24]:
```
# Replacing all castes with numerical values

df['Caste']=df['Caste'].replace(['OC','BC-A','BC-C','BC-D','BC-B','BC-E','SC','ST'],[0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8])
```

IMPLEMENTATION

```
# Replacing Gender with numerical values

df['Gender']=df['Gender'].replace(['Male','Female'],[0.34,0.66])
```

```
df.head()
```

	Hall ticket (10th class)	10th GPA	Govt/Private School	Caste	Gender	Mandal	District	State	Chance of Admit
0	1720104115	10.0	0.0	0.5	0.34	Kallur	Kurnool	Ap	0.83
1	1722136839	10.0	0.4	0.5	0.34	Galiveedu	Kadapa	Ap	0.98
2	1718124014	9.7	0.4	0.5	0.66	Madanapalli	Chittoor	Ap	0.93
3	1722136791	9.8	0.4	0.7	0.34	Galiveedu	Kadapa	Ap	0.95

IMPLEMENTATION

In [28]:
```
#Calculating performance score with major factors which affects chance of admit

df["Performance Score"] = df["Govt/Private School"] + df["Caste"] + df["Gender"] + df["10th GPA"]
```

In [29]:
```
df.head()
```

Out[29]:

	Hall ticket (10th class)	10th GPA	Govt/Private School	Caste	Gender	Mandal	District	State	Chance of Admit	Performance Score
0	1720104115	10.0	0.0	0.5	0.34	Kallur	Kurnool	Ap	0.83	10.84
1	1722136839	10.0	0.4	0.5	0.34	Galiveedu	Kadapa	Ap	0.98	11.24
2	1718124014	9.7	0.4	0.5	0.66	Madanapalli	Chittoor	Ap	0.93	11.26
3	1722136791	9.8	0.4	0.7	0.34	Galiveedu	Kadapa	Ap	0.95	11.24
4	1722107642	9.8	0.4	0.1	0.66	Vallur	Kadapa	Ap	0.90	10.96

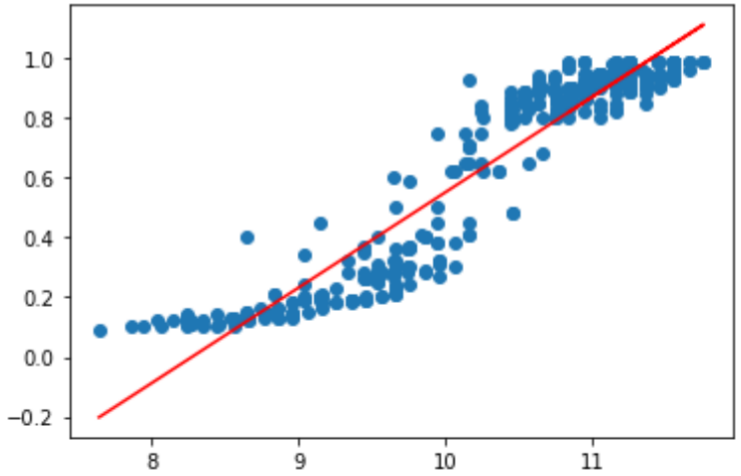
In [30]:
```
X=df["Performance Score"].values
```

In [31]:
```
Y=df["Chance of Admit"].values
```

IMPLEMENTATION

In [33]:
```
a, b = np.polyfit(X, Y, 1)
plt.scatter(X,Y)
plt.plot(X,a*X+b,"r")
```

Out[33]:
```
[<matplotlib.lines.Line2D at 0x189bdfcca90>]
```



IMPLEMENTATION

```
x=df.loc[:,["10th GPA","Govt/Private School","Caste","Gender"]] # consedering 4 columns that chance of admit depended on
```

```
y=df["Chance of Admit"]
```

```
#Dividing dataset into train and test data

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=45)
```

In [84]:
```
# Linear Regression

from sklearn.linear_model import LinearRegression
LR=LinearRegression()
LR.fit(x_train,y_train)
y_pred=LR.predict(x_test)
```

In [85]:
```
LR.predict([[9.8,0.4,0.6,0.34]])
```

Out[85]:
```
array([0.93105779])
```

IMPLEMENTATION

In [45]:
```
#Calculating accuracy and error of model for this project

from sklearn.metrics import mean_absolute_error,r2_score
print("R2 score of the model is ",r2_score(y_pred,y_test))
print("mean_absolute_error  of the model is ",mean_absolute_error(y_pred,y_test))
```

```
R2 score of the model is  0.9417064974981846
mean_absolute_error  of the model is  0.05599397966866188
```

In [46]:
```
#Predicting the admission percentage

LR.predict([[9.7,0.4,0.7,0.34]]) # gpa , school_type , caste , gender
```

Out[46]:
```
array([0.9020695])
```

In [47]:
```
print(LR.coef_)
```

```
[ 0.34405909  0.25141407  0.05417621 -0.00192658]
```

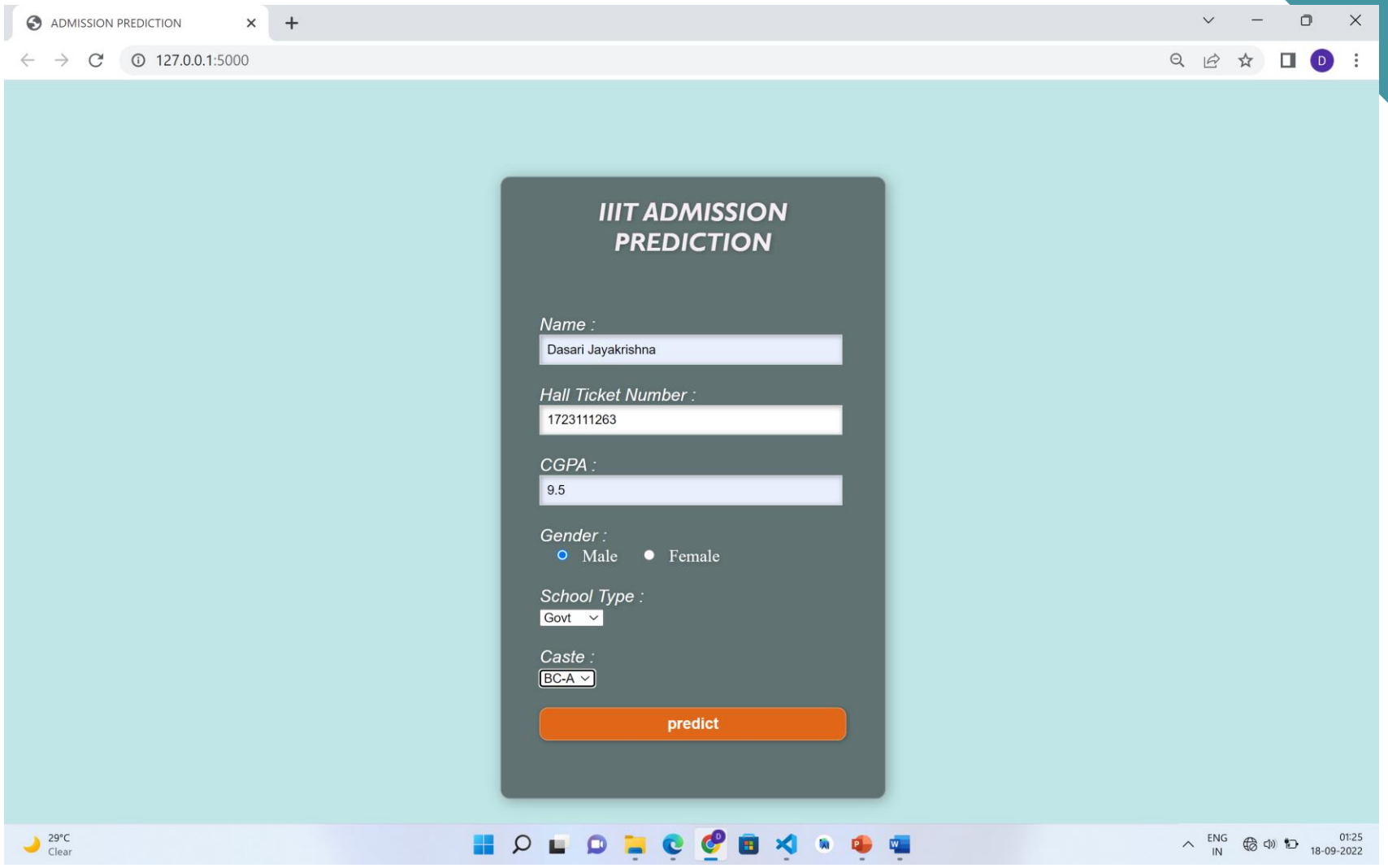
In [48]:
```
print(LR.intercept_)
```

```
-2.5731376452650485
```

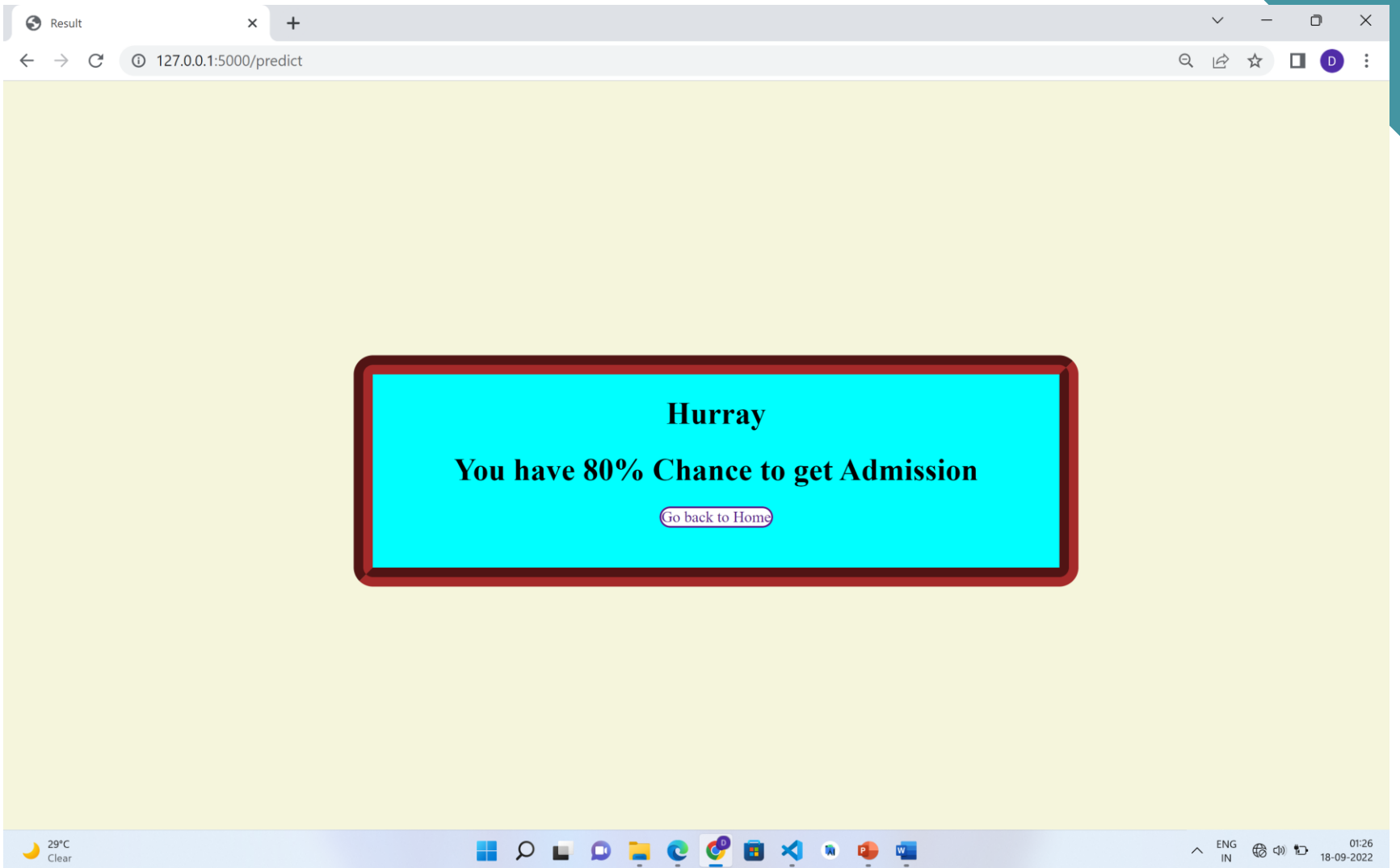
TECHNOLOGIES

- FRONTEND : HTML, CSS.
- BACKEND : Flask.
- ENVIRONMENT : Jupyter Notebook, Visual Studio Code.

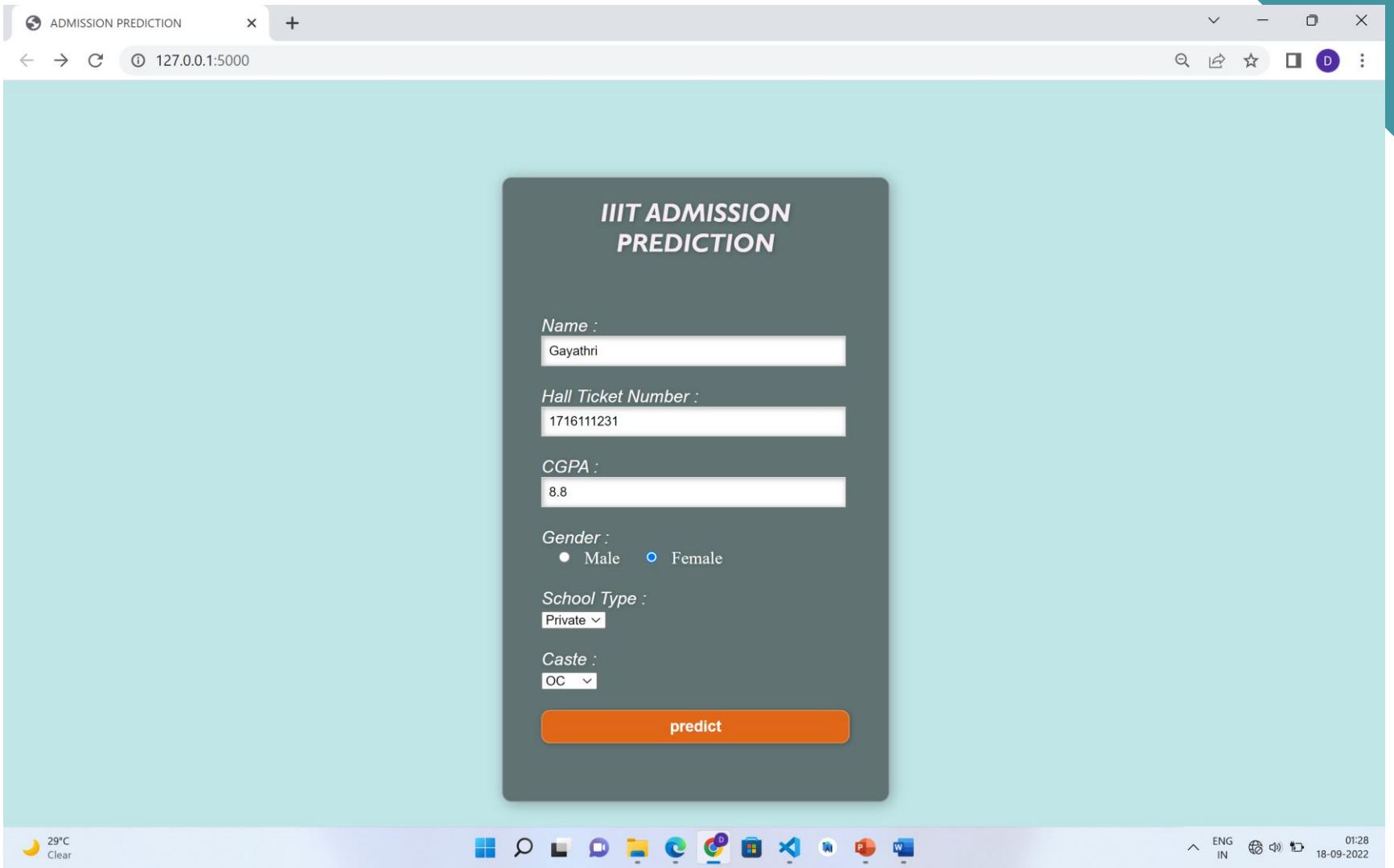
RESULTS


ADMISSION PREDICTION
127.0.0.1:5000
IIIT ADMISSION PREDICTION
Name :
Dasari Jayakrishna
Hall Ticket Number :
1723111263
CGPA :
9.5
Gender :
Male
Female
School Type :
Govt
Caste :
BC-A
predict
29°C
Clear
ENG
IN
01:25
18-09-2022

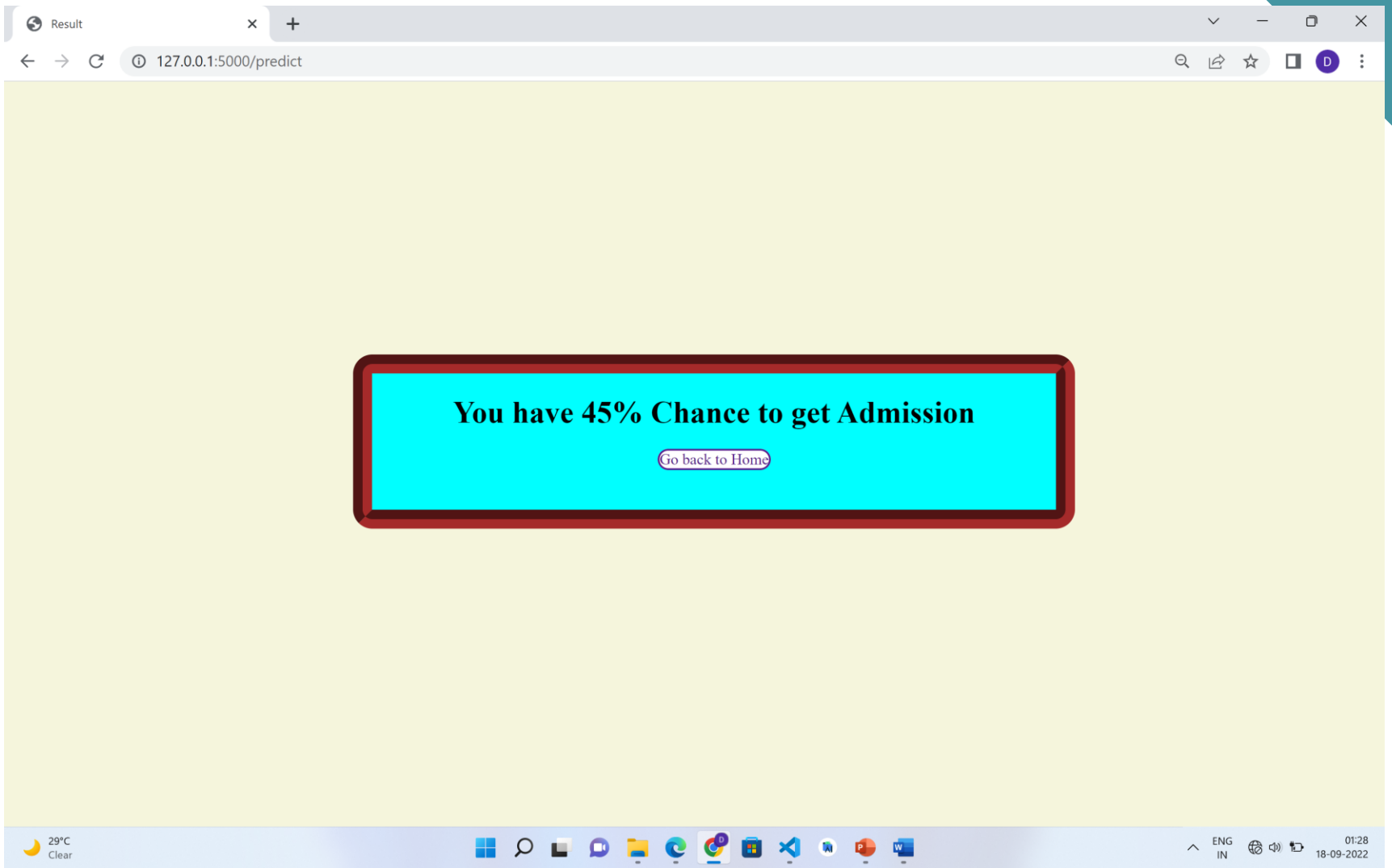
RESULTS


Result
127.0.0.1:5000/predict
Hurray
You have 80% Chance to get Admission
Go back to Home
29°C
Clear
ENG
IN
01:26
18-09-2022

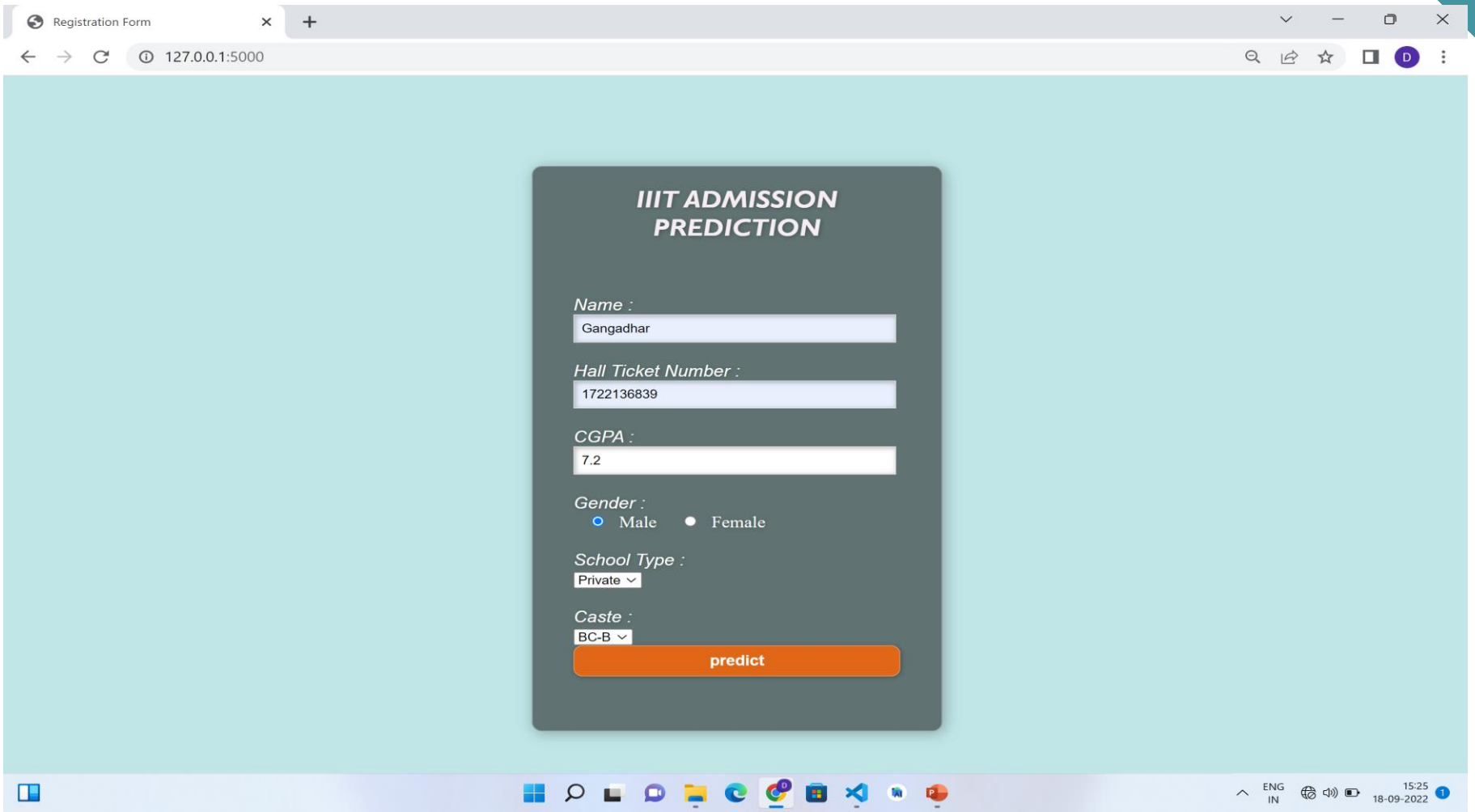
RESULTS


ADMISSION PREDICTION
127.0.0.1:5000
IIIT ADMISSION PREDICTION
Name :
Gayathri
Hall Ticket Number :
1716111231
CGPA :
8.8
Gender :
Male
Female
School Type :
Private
Caste :
OC
predict
29°C
Clear
ENG
IN
01:28
18-09-2022

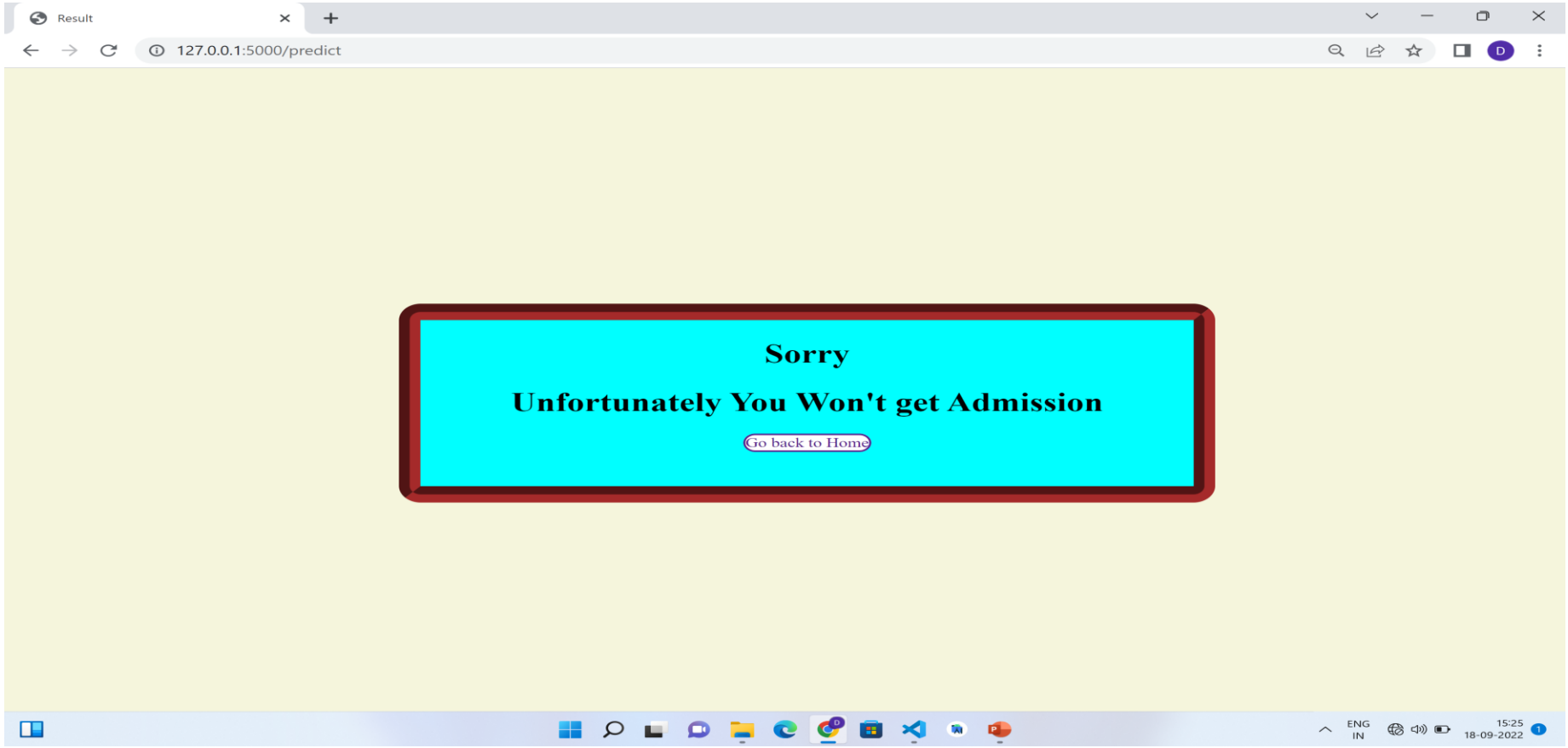
RESULTS


Result
127.0.0.1:5000/predict
You have 45% Chance to get Admission
Go back to Home
29°C
Clear
ENG
IN
01:28
18-09-2022

RESULTS


Registration Form
127.0.0.1:5000
IIIT ADMISSION PREDICTION
Name :
Gangadhar
Hall Ticket Number :
1722136839
CGPA :
7.2
Gender :
Male
Female
School Type :
Private
Caste :
BC-B
predict

RESULTS


Result
127.0.0.1:5000/predict
Sorry
Unfortunately You Won't get Admission
Go back to Home

CONCLUSION

- This predictor system provides students to early predict their admission.
- As the model built based on machine learning, it learns the prediction based on past data. Whenever student enter his details the model gives the prediction based on similar previous data.
- Students can estimate the percentage of chance of admission , so that they can think about further ways of their admission.

FUTURE ENHANCEMENTS

- We will be considering option of special candidates like NCC,CAP,PH etc. There will be more priority for these special candidates, so automatically chance of admission for these candidates increases.

- We will also try to improve the interface more interactive i.e. queries section, help section etc.

REFERENCES

- GEEKS FOR GEEKS : https://www.geeksforgeeks.org/ml-linear-regression/
- Scikit Learn : https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- Javatpoint : https://www.javatpoint.com/machine-learning
- ResearchGate : https://www.researchgate.net/publication/348433004/
- YouTube

THANK YOU