TAMIL NADU ENGINEERING ADMISSIONS PREDICTOR

SEMINAR-II PROJECT REPORT

Submitted by

A P AISHWARYA LAKSHMI - RA2011026020066

INDHUMATHI S - RA2011026020088

RIDHANYA P - RA2011026020104

Under the guidance of

Mr. T. RAJESH

Dr. T.P. LATCHOUMI

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM CAMPUS, CHENNAI-600089

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/S 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that the Seminar-II report titled "**TAMIL NADU ENGINEERING ADMISSIONS PREDICTOR**" the bonafide work of **A P AISHWARYA LAKSHMI [RA2011026020066], INDHUMATHI S [RA2011026020088], RIDHANYA P [RA2011026020104]** submitted for the course 18CSP106L Seminar –II . This report is a record of successful completion of the specified course evaluated based on literature reviews and the supervisor. No part of the Seminar Report has been submitted for any degree, diploma, title, or recognition before.

SIGNATURE

SUPERVISOR
Mr. T Rajesh, M. Tech.,
Assistant Professor,
Department of CSE,
RamapuramSRM IST, Ramapuram

SIGNATURE

HEAD OF THE DEPARTMENT
Dr.K.Raja, M.E., Ph.D
Department of CSE,
SRM IST,

Submitted for the Seminar-II Viva Voce Examination held on _____ at SRM Institute of Science and Technology, Ramapuram Campus, Chennai-600089.

EXAMINER 1 **EXAMINER 2**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY,RAMAPURAM, CHENNAI – 89

DECLARATION

We hereby declare that the entire work contained in this project report titled **"TAMIL NADU ENGINEERING ADMISSIONS PREDICTOR"** has been carried out by **A P AISHWARYA LAKSHMI [RA2011026020066], INDHUMATHI S [RA2011026020088], RIDHANYA P [RA2011026020104]** at SRM Institute of Science and Technology, Ramapuram Campus, Chennai-600089, under the guidance of Mr.T.Rajesh M.Tech., Assistant Professor, Department of Computer Science and Engineering.

Place:
Chennai
Date:

A P AISHWARYA LAKSHMI	INDHUMATHI S	RIDHANYA P
RA2011026020066	RA2011026020088	RA2011026020104

ABSTRACT

Increasing student admission and enrollment, especially in engineering and computing programs, is a desirable goal for many universities. At the same time, this goal can be difficult to achieve. The aim of this project is to develop a data analytics model that can be used by universities and colleges to improve student admission and enrollment process.

Predictive analytics is the technique of using historical data to create, test, and validate a model to best describe and predict the probability of an outcome.In India, we have a huge crowd of students writing their 12th board exams every year.

And the college admission process is a whole new thing. Students even start preparing for IIT exams right from their 8th standards. And we have various institutes training children just to get them into top colleges like IITs, and MITs. There are so many ways to get into colleges, though. In Tamilnadu, we have the TNEA process.

We also have centralized IIT exams for all students from India. To make any process easier, visualization and proper understanding of data is needed, or the ability to predict. There has also been an increasing demand for courses like Computer Science, especially Big Data, IoT, Artificial Intelligence, Cloud computing, and such.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The purpose of this project is to predict college admissions in Tamil Nadu. Applying to college can feel uncertain and even mysterious. Students and supporters are inundated with information through marketing and media, and it quickly becomes an overwhelming experience. It doesn't have to be. In its most perfect form, the college search should be about making a good match between one's strengths and interests and a school that will support their aspirations and growth. For this to happen, students need better, more accessible, data.

A college admissions predictor is a tool that uses statistical models and data analysis to predict a student's chances of being admitted to a particular college or university. It takes into account various factors that are typically considered by admissions officers, such as a student's academic performance, test scores, extracurricular activities, essays, and recommendations.

Using a college admissions predictor can help students and their families make more informed decisions about which colleges to apply to, and how to allocate their time and resources in the application process. By inputting their credentials and background information, students can receive an estimate of their likelihood of being admitted to a particular school, and can use this information to prioritize their applications and set realistic expectations.

It is important to note, however, that college admissions predictors are not perfect, and should be used as just one tool in the college search and application process. Admissions decisions are often influenced by subjective factors that cannot be easily quantified, and each school has its own unique admissions process and criteria. Nevertheless, a college admissions predictor can be a useful starting point for students as they navigate the complex and competitive world of college admissions.

2. PROBLEM STATEMENT

The college admissions process is often stressful and overwhelming for students and their families, with many factors to consider and a high level of uncertainty about the outcome. Even with a strong academic record and impressive extracurricular achievements, it can be difficult to predict one's chances of being admitted to a particular college or university.

This is where a college admissions predictor can be helpful, by providing students with an estimate of their likelihood of being admitted to a given school based on various factors. However, there are several challenges associated with developing an accurate and reliable college admissions predictor.

First, there is a vast amount of data that needs to be collected and analyzed, including information about students' academic performance, test scores, and other factors. This data must be accurate and up-to-date, and must be analyzed using sophisticated statistical models and machine learning algorithms.

Second, there are many subjective factors that can influence admissions decisions, such as their community, the strength of their recommendations, reservations and such. These factors can be difficult to quantify and incorporate into a predictive model.

Addressing these challenges will require careful data collection and analysis, as well as ongoing refinement and improvement of predictive models based on feedback from students and admissions officers. Ultimately, a successful college admissions predictor will be one that is accurate, reliable, and transparent, and that helps students make informed decisions about their college applications.

3. SCOPE AND OBJECTIVE

Objective :

The aim of the project is to predict the college that a student will be getting into with his PCM marks and his community.

Project Domain: The domain of the project is Machine Learning. We have used Supervised Learning techniques to obtain the desired results.

Scope :

The scope of a college admissions predictor can vary depending on its design and intended audience. At a basic level, a predictor can provide a simple estimate of a student's chances of being admitted to a particular college or university based on their academic credentials and other factors. More advanced predictors may incorporate additional data points, such as information about a student's extracurricular activities, essays, and recommendations, in order to provide a more accurate prediction.

The scope of a college admissions predictor may also depend on the institutions it is designed to predict admissions outcomes for. Some predictors may focus on a specific set of colleges or universities, while others may aim to provide predictions for a wider range of institutions. Additionally, predictors may be designed for different types of students, such as high school seniors or transfer students, and may take into account different factors depending on the student population.

Another aspect of the scope of a college admissions predictor is its level of transparency and accessibility. A good predictor should be transparent about the data and models used to make predictions, and should provide clear explanations of how the predictions are generated.

The predictor should also be accessible to all students, regardless of their background or resources, and should be designed in a way that is easy to use and understand.

Overall, the scope of a college admissions predictor should be tailored to the needs of its users, and should aim to provide accurate, reliable, and accessible predictions that can help students make informed decisions about their college applications.

4. EXISTING SYSTEM

Traditionally, institutions have advertised themselves by posting information on their websites and using multimedia. However, these traditional methods are increasingly becoming insufficient on their own .

Therefore, they should be supported by a predictive analytics approach that utilizes personal attributes to appeal to the interests of prospective students; and predict the probability that students will accept an offer and enroll into a course.

More so, predictive analytics can provide accurate information and knowledge about future admission trends, and thus support planning, resource allocation, and decision making regarding the growth of an institution. If the school has an anticipation for a growth in student enrollment then they can plan accordingly to provide adequate resources required to educate students.

5. LITERATURE SURVEY

Sl. No.	Journal Name	Year of Publishing	Paper Title	Author	Description
1.	Using Technology in Undergraduate Admission	2008	Technology in Undergraduate Admission	Robin Lindbeck and Brian Fodrey	To identify the current practices and future plans for using technology in admission practices at four year colleges and universities.
2.	Research by Heather Levesque	2014	Gen Z Students' Experiences with College Choice	Heather Levesque-(East Tennessee State University)	Data mining approach has been used used, and the investigation is narrowed down and could have a higher scope for larger population.
3.	Research by Baiou and Balinsk Team	2015	A graphical modeling of student admissions and faculty recruitment problems	Baiou and Balinsk	A study and analysis of admission data and student educational outcomes is presented in Heinesen

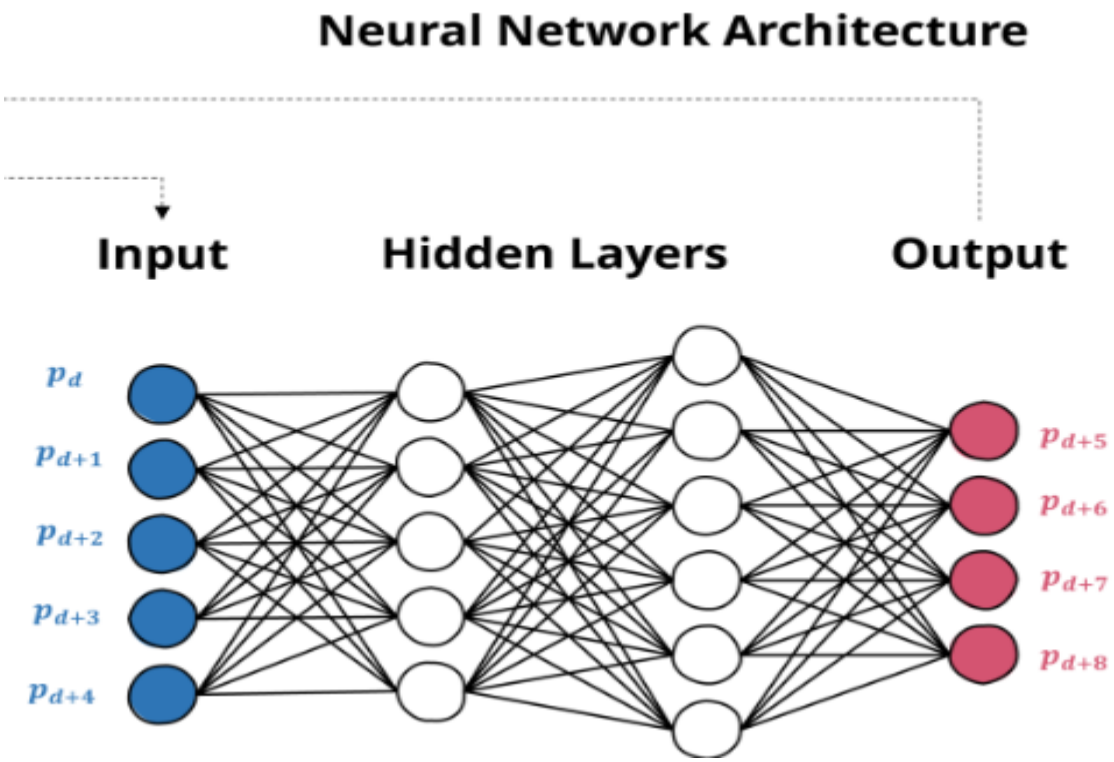
4.	Research done by Ching-Ling Wu and Haiyan Bai	2016	From early aspirations to actual attainment: the effects of economic status and educational expectations on university pursuit.	Ching-Ling Wu and Haiyan Bai	This study investigated the effects of economic status and the educational expectations of significant others on early university aspirations and actual university attainment.
5.	Research on time series data prediction based on clustering algorithm by Yaebau	2014	Research on time series data prediction based on clustering algorithm	Yaebau	Based on this situation, this paper analyzed the data of Yuebao, and according to the user's attributes and the operating characteristics, this paper classified 567 users of Yuebao, and made further predicted the data of Yuebao for every class of users, the results showed that the forecasting model in this paper can meet the demand of forecasting.

6.	Research based on Mining Time Series Data with AprioriTid Algorithm	2014	Mining Time Series Data with AprioriTid Algorithm	Hiran Kumar Deva Sharma, Swapnil Mishra	The performances of both the algorithms in terms of computation time requirements for generating frequent itemsets are analyzed and corresponding corrections are to be made.
7.	ARIMA Time Series Application	2015	ARIMA Time Series Application to Employment Forecasting	Xiaoguo Wang, and Yuejing Liu	The paper establishes an ARIMA model on the employment information of the computer industry from 2002 to 2007 in China, and using the model, gives a prediction of the situation in 2008.

8.	Performance comparison and future estimation of time series data using predictive data mining techniques	2017	Performance comparison and future estimation of time series data using predictive data mining techniques	\Harshita Tanwar; Misha Kakkar	These two models are analyzed with respect to standardized error generated by them, fitted values, residuals, standardized error, square root standardized error are used for forecasting future expenditure prediction. Result shows that both models accurately and approximately predict the same future Expenditure measure.

9.	Decision Tree Classification and Forecasting of Pricing Time Series Data	2016	Decision Tree Classification and Forecasting of Pricing Time Series Data	Emil Lundkvist	Although the tool described in this thesis is already distributed in the company and works as it was intended, there are of course many improvements that can be made. The most important ones are described in the following sections.
10.	Research by Robin Lindbeck and Brian Fodrey	2016	Using Technology in Undergraduate Admission: A Student Perspective	Robin Lindbeck and Brian Fodrey	This inquiry offers several opportunities for additional research. First, a small geographically homogenous convenience sample was used in this inquiry.

6. ARCHITECTURE



7. PROPOSED WORK

- The idea is to build a College admissions predictor.
- A college admissions predictor would be a valuable tool for high school students, who are often unsure of their chances of admission to different universities. Such a predictor could consider factors such as PCM, caste, and so on and use machine learning algorithms to make predictions based on historical data.
- A college admissions predictor can be created using machine learning algorithms that analyze a variety of data points. Here's a list of data points that are included in the predictor:

1. Physics, Chemistry and Mathematics (PCM)
2. College name
3. Branch name
4. Demographic information (Caste)

- Once these data points are collected and organized, model is trained using machine learning algorithms to predict a student's likelihood of being admitted to certain colleges or universities.
- The predictor can be created by training a model on historical admissions data, where the model learns from the patterns and trends in past applicants that were accepted or rejected from different colleges. This model can then be used to analyze new applicants and determine their chances of admission.
- This predictor can be made available as a free online tool for high school students to use. However, it's important to note that college admissions are a complex process, and other factors may also influence admissions decisions.

ALGORITHM

The steps involved in building proposed model:

1. Data Collection: Collect a large dataset that includes historical admission data, including student profiles such as PCM marks, and additional contextual factors such as race (caste).
2. Data Preprocessing: Clean and process the data by removing duplicates, identifying missing values, and encoding categorical variables.
3. Feature Selection: Identify the key predictors that could influence the admission decision.
4. Model Building: Apply machine learning algorithms and techniques (such as decision trees, support vector machines, neural networks, regression models, etc.) to predict the likelihood of a student getting admission.
5. Model Evaluation: Using a set of historical data, evaluate the model's accuracy and adjust it based on the feedback.
6. Deployment: Once the model is refined and accurate, deploy it using streamlit to provide personalized recommendations to incoming applicants based on their academic profiles.
7. Continuous Improvement: As new data becomes available, continue to refine the model to improve its accuracy and ability to predict admission outcomes.

8. FUTURE SCOPE

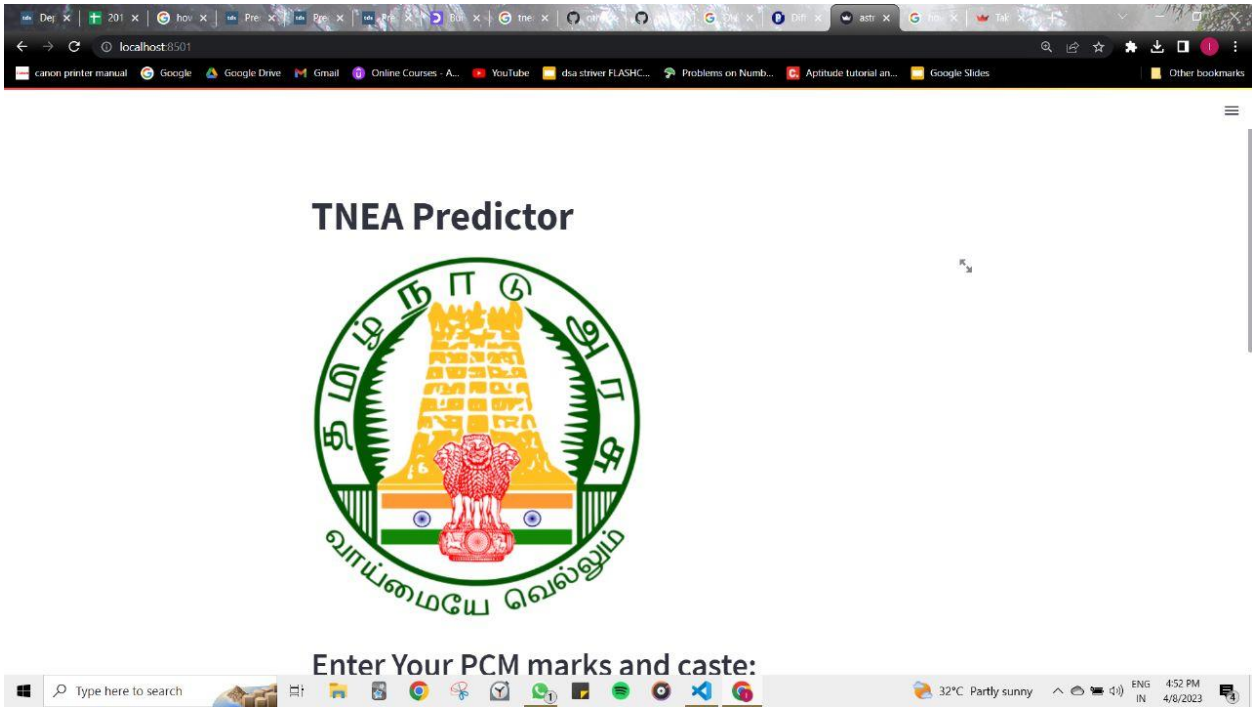
Some points to consider are:

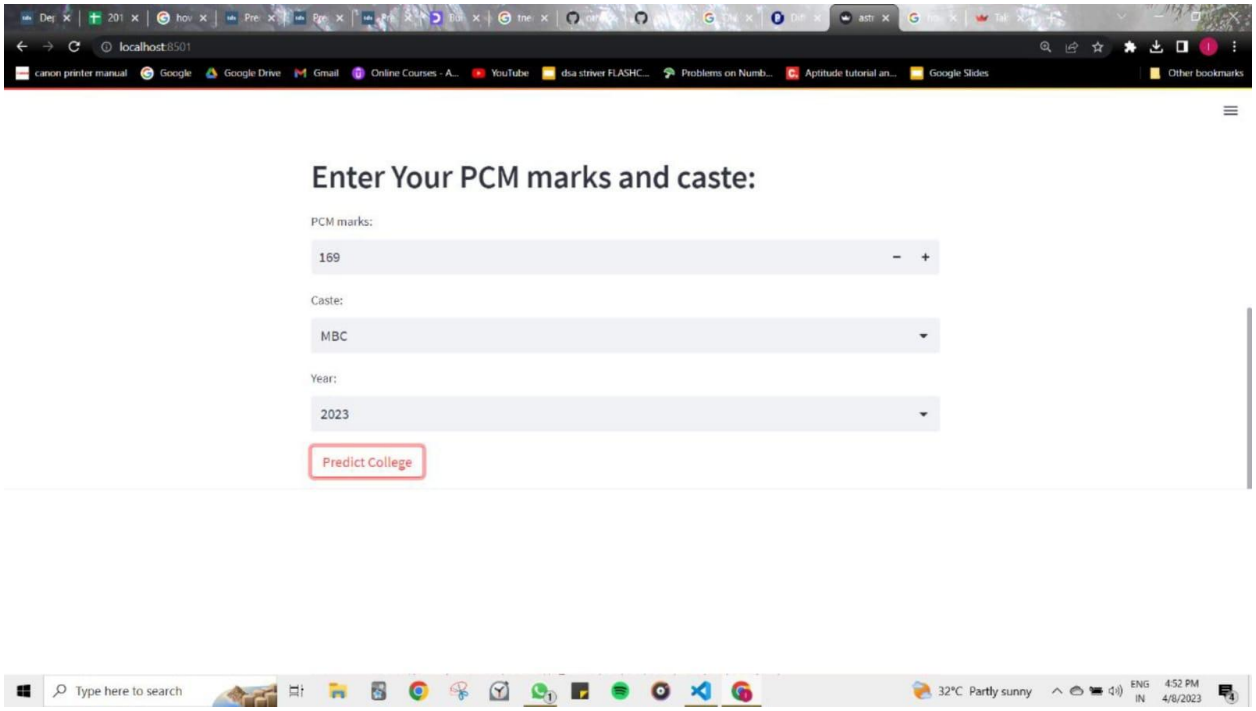
1. Personalized Recommendations: With the advancements in data science, machine learning and AI, college admission predictors have the potential to offer personalized recommendations for students based on their academic background, interests, and career aspirations.
2. Improved Accuracy: College admission predictors can leverage historical data to improve their accuracy. This will help students and parents make more informed decisions when it comes to college admissions.
3. Enhanced Accessibility: College admission predictors can be made easily accessible to a larger audience, including students from low-income families or those who come from rural areas where resources for educational planning may be limited.
4. Usefulness for Admissions Officers: College admission predictors can be helpful for admissions officers as well. By analyzing historical data, admission officers can identify trends and patterns that can help them make more informed decisions on who to admit.
5. Worldwide Reach: College admission predictors can be made available worldwide, allowing students around the world to access information about colleges and universities that they may not have known about otherwise.
6. Time Saving Solutions: College admission predictors save time and offer more comfort to the student, as they do not have to search for information for many hours.Overall, college admission predictors have a bright future and have the potential to revolutionize the college admissions process.

9. DISADVANTAGES

1. Overconfidence: Relying heavily on an admission predictor can create unnecessary overconfidence in a student's application.
2. Financial Cost: Some college admission predictors may charge a fee for access to the service. These fees can be costly and may not be necessary, especially if the tool's accuracy is questionable.
3. Inefficient Use of Time: Using a college admission predictor can be a significant time investment.
4. Stress-inducing: The use of college admission predictors may cause unnecessary stress and anxiety for students and parents, especially if the prediction is negative. This can affect student morale and self-confidence, leading to a negative impact on the application overall.

10. IMPLEMENTATION SNAPSHOTS





11. IMPLEMENTATION CODE

In [22]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [23]:
```
import pandas as pd
df = pd.read_csv("C:/Users/indhu/OneDrive/Desktop/tnea-proj/data/2017-2020.csv")
```

In [24]:
```
df.head()
```

Out[24]:

	Year	College Code	College_Name	Branch Code	Branch Name	OC	BC	BCM	MBC	SC	SCA	ST
0	2017	1	UNIVERSITY DEPARTMENTS OF ANNA UNIVERSITY, CHE...	AI	Agricultural and Irrigation Engg.(SS)	196.25	195.25	193.25	194.25	188.75	185.75	175.25
1	2017	1	UNIVERSITY DEPARTMENTS OF ANNA UNIVERSITY, CHE...	BY	Bio Medical Engg(SS)	198.25	197.50	197.25	196.50	193.00	191.75	188.00
2	2017	1	UNIVERSITY DEPARTMENTS OF ANNA UNIVERSITY, CHE...	CE	Civil Engineering	198.50	198.00	197.50	197.50	196.00	193.50	196.00
3	2017	1	UNIVERSITY DEPARTMENTS OF ANNA UNIVERSITY, CHE...	CM	Computer Science and Engg. (SS)	199.00	198.75	199.00	197.75	193.50	190.50	193.33
4	2017	1	UNIVERSITY DEPARTMENTS OF ANNA UNIVERSITY, CHE...	CS	Computer Science and Engg.	199.75	199.50	199.50	199.00	197.50	196.25	193.75

In [25]:
```
castes = list(df.columns[5:])
castes
```

Out[25]:
```
['OC', 'BC', 'BCM', 'MBC', 'SC', 'SCA', 'ST']
```

In [26]:
```
dic = {
   'Year' : [],
    'College Code' : [],
    'Branch Code' : [],
    'Caste':[],
    'Mark': []
}

new_df = pd.DataFrame(dic)
new_df
```

Out[26]:

Year	College Code	Branch Code	Caste	Mark

In [27]:
```
for index, row in df.iterrows():
    for caste in castes:
        new_df.loc[len(new_df.index)] = [row["Year"],
                                         row["College Code"],
                                         row["Branch Code"],
                                          caste,
                                          row[caste]
                                         ]
```

In [28]:
```
new_df
```

Out[28]:

	Year	College Code	Branch Code	Caste	Mark
0	2017	1	AI	OC	196.25
1	2017	1	AI	BC	195.25
2	2017	1	AI	BCM	193.25
3	2017	1	AI	MBC	194.25
4	2017	1	AI	SC	188.75
...	...	...	...	...	...
1675	2020	1014	EC	BCM	86.00
1676	2020	1014	EC	MBC	86.00
1677	2020	1014	EC	SC	81.00
1678	2020	1014	EC	SCA	NaN
1679	2020	1014	EC	ST	NaN

1680 rows × 5 columns

In [29]:
```
new_df.to_csv("C:/Users/indhu/OneDrive/Desktop/tnea-proj/data/final_data.csv", index=False)
```

In [30]:
```
new_df = pd.read_csv("C:/Users/indhu/OneDrive/Desktop/tnea-proj/data/final_data.csv")
```

In [43]:

```
mapping_branch = {}
branch = list(new_df["Branch Code"].unique())

for x in range(len(branch)):
    mapping_branch[branch[x]] = x

mapping_branch
```

Out[43]:

```
{0: 0,
 1: 1,
 2: 2,
 3: 3,
 4: 4,
 5: 5,
 6: 6,
 7: 7,
 8: 8,
 9: 9,
 10: 10,
 11: 11,
 12: 12,
 13: 13,
 14: 14,
 15: 15,
 16: 16,
 17: 17,
 18: 18,
 19: 19,
 20: 20,
 21: 21,
 22: 22,
 23: 23,
 24: 24,
 25: 25,
 26: 26,
 27: 27,
 28: 28,
 29: 29,
 30: 30,
 31: 31,
 32: 32,
 33: 33,
 34: 34,
 35: 35,
 36: 36,
 37: 37,
 38: 38,
 39: 39,
 40: 40,
 41: 41}
```

In [44]:
```
mapping_caste = {}
caste = list(new_df["Caste"].unique())

for x in range(len(caste)):
    mapping_caste[caste[x]] = x

mapping_caste
```

Out[44]:
```
{0: 0, 1: 1, 2: 2, 3: 3, 4: 4, 5: 5, 6: 6}
```

In [45]:
```
def cat2num_encoding(label, mapping):
    # integer representation
    for x in range(len(label)):
        label[x] = mapping[label[x]]
    return label
```

In [46]:
```
new_df["Caste"] = cat2num_encoding(new_df["Caste"], mapping_caste)
new_df["Branch Code"] = cat2num_encoding(new_df["Branch Code"], mapping_branch)
```

In [47]:
```
new_df.head()
```

Out[47]:

	Year	College Code	Branch Code	Caste	Mark
0	2017	1	0	0	196.25
1	2017	1	0	1	195.25
2	2017	1	0	2	193.25
3	2017	1	0	3	194.25
4	2017	1	0	4	188.75

In [48]:
```
from sklearn.multioutput import MultiOutputClassifier
from sklearn.ensemble import RandomForestClassifier
```

In [51]:
```
for label, content in new_df.items():
    if pd.api.types.is_numeric_dtype(content):
        if pd.isnull(content).sum():
            #df_tmp[label+"_is_missing"] = pd.isnull(content)
            new_df[label] = content.fillna(content.median())
```

In [52]:
```
X = new_df.drop(columns = ["College Code", "Branch Code"])
Y = new_df.drop(columns = ["Year", "Caste", "Mark"])
```

In [53]:
```
X.head()
```

Out[53]:

	Year	Caste	Mark
0	2017	0	196.25
1	2017	1	195.25
2	2017	2	193.25
3	2017	3	194.25
4	2017	4	188.75

In [54]:
```
Y.head()
```

Out[54]:

	College Code	Branch Code
0	1	0
1	1	0
2	1	0
3	1	0
4	1	0

In [55]:
```
Y.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1680 entries, 0 to 1679
Data columns (total 2 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   College Code  1680 non-null   int64 
 1   Branch Code   1680 non-null   object
dtypes: int64(1), object(1)
memory usage: 26.4+ KB
```

In [56]:
```
Y["Branch Code"] =Y["Branch Code"].astype('int')
```

In [57]:
```
clf = MultiOutputClassifier(RandomForestClassifier()).fit(X, Y)
```

In [58]:
```
clf.predict(X[-2:])
```

Out[58]:
```
array([[5, 5],
       [5, 4]], dtype=int64)
```

```python
import xgboost as xgb
import streamlit as st
import pandas as pd

#Loading up the Regression model we created
model = xgb.XGBRegressor()
model.load_model('axgb_model.json')

#Caching the model for faster loading
@st.cache

def predict(Year,College_Code, College_Name, Branch_Code,Branch_Name,OC,BC,BCM,MBC,SC,SCA,ST):

    prediction = model.predict(pd.DataFrame([[Year,College_Code, College_Name, Branch_Code,Branch_Name,OC,BC,BCM,MBC,SC,SCA,ST]], columns=['Year','College_Code', 'College_Name
    return prediction


st.title('TNEA Predictor')
st.image("images\logo.png")
st.header('Enter Your PCM marks and caste:')



marks = st.number_input('PCM marks:', min_value=0, max_value=200, value=1)
caste = st.selectbox('Caste:', ['OC','BC','BCM','MBC','SC','SCA','ST'])
Year = st.selectbox('Year:', ['2017','2018','2019','2020','2021','2022','2023'])

if st.button('Predict College'):
    college = predict(College_Name,OC,BC,BCM,MBC,SC,SCA,ST)
    st.success(f'The predicted College is ${college[0]}')
```

12. CONCLUSION

In conclusion, a college admission predictor can be a helpful tool for students and families as they navigate the college application process. By offering estimates of a student's likelihood of being admitted to particular schools, the predictor can provide guidance on where to focus one's efforts and how to allocate resources. However, it is important to remember that these predictors are not infallible and should be used in conjunction with other sources of information and advice. It is also essential to recognize that college admission predictors are based on historical data, and the decisions of admission committees can be influenced by many unforeseen factors. Therefore, students should not rely solely on these predictors but should approach the application process with an open mind, a positive attitude, and a willingness to take risks. Ultimately, success in college admissions will depend on a combination of hard work, perseverance, self-reflection, and a bit of luck. By using college admission predictors thoughtfully and in combination with other resources, students and families can make informed decisions about where to apply and increase their chances of finding the right fit for their academic, social, and personal goals. Ultimately, the decision of which college to attend is complex and multifaceted, taking into account a range of factors such as academic fit, financial aid, location, and personal preferences.