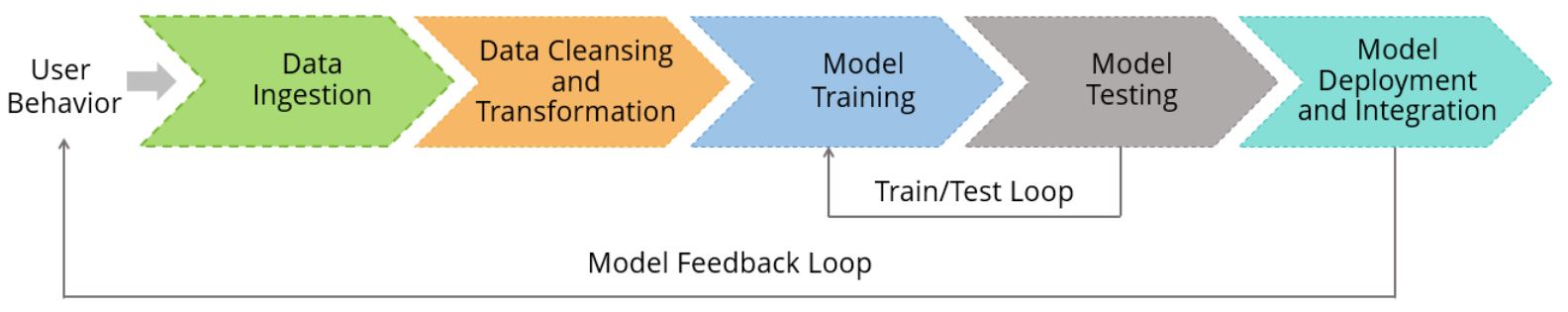
UNIVERSITY ADMISSION PREDICTION

**1.1Indtroduction:**

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**. Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.



*Some machine learning methods*

Machine learning algorithms are often categorized as supervised or unsupervised.

#### **Supervised machine learning** algorithmscan apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, ***Unsupervised machine learning*** algorithmsare used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

***Reinforcement machine learning*** algorithmsis a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

***1.2. Objective of Research:***

A persons education plays a vital role in their life. While planning for education students often have several questions regarding the courses, universities, job opportunities, expenses involved, etc. Securing admission in their dream university is one of their main concerns. It is seen that often students prefer to pursue their education from universities which have global recognition. And when it comes to international students the ﬁrst choice of the majority of them is the United States of America. With the majority of worlds highly reputed universities, wide range of courses oﬀered in every sector, highly accredited education system and teaching, scholarships provided to students, best job market and many more advantages make it the dream destination for the international students. According to research, there are above 8 Million international students studying in more than 1700 public and 2500 private universities and colleges across the USA. (MasterPortal (2017)) The majority of international students studying in the USA are from India and China. In the past decade, India has seen a huge increase in the number of students opting to pursue their education from foreign universities in countries like The USA, Ireland, Australia, Germany, etc. Although there are signiﬁcant universities and colleges in India, students are ﬁnding it diﬃcult to get admission in the highly ranked colleges and also getting a job is a challenge as the ratio of number students to the number work opportunities available is quite high. India is one of the leading counties in the number of software engineers produced each year; it becomes tough for the students to ﬁnd jobs in elite companies due to high competition. This motivates a good number of students to pursue post-graduation in their ﬁeld. It is seen that the number of students pursuing Masters in Computer Science ﬁeld from universities in the USA is quite high; the focus of this research will be on these students. (Times (2015)) Majority of universities in the USA follow similar guidelines for providing admission to students. Universities take into consideration diﬀerent factors like score on aptitude based examination like the General Record Examination (GRE), command over the English language is judged based on their score in English competency test like Test Of English as a Foreign Language (TOEFL) OR International English Language Testing System (IELTS), their work experience in same or other ﬁelds, the quality of the Letters Of Recommendation (LOR) and the Statement Of Purpose documents provided by the student etc. Based on the overall proﬁle of the student decision is taken by the universities admission team to admit or reject a particular candidate.

Every candidate has to take all the required examination and build a strong proﬁle to secure admission in their dream universities in the USA. Once the candidates have made their proﬁle ready, they apply to the universities where they aim to secure admission. The students have to shortlist the universities which are best known for the courses they are looking for and also they should have an idea about their chances of securing admission in those universities based on their proﬁle. This task of shortlisting the universities where the student has high chances of admission is diﬃcult for mainly for the international students, so they end up with applying to many universities in hopes of getting admission in few of them thus investing an extra amount of money in the applications. There are several portals and websites which provide information and help to students in shortlisting the universities, but they are not reliable. Most of the students dont take the risk of evaluating the colleges by themselves, and they seek the help of the education consultancy ﬁrms to do it for them. Again for this students have to pay a huge amount of fee to the education consultant.

The primary objective of this research is to develop a system to solve the problems the international students are facing while applying for universities in the USA. We will be developing a Student Admission Predictor (SAP) system which will help the students to predict the chances of their application being selected for a particular university for which they wish to apply based on their proﬁle. Also, the system will provide a recommendation of universities to the student to which the student has a high possibility of getting admission. Multiple machine learning classiﬁcation algorithms were evaluated to develop the system.

Our goal is to apply machine learning algorithms to admission data set. University Selection model is used by the students to find the probability of the student to get an admit in the university before applying.

***1.3. Problem Statement:***

Problem Statement is to find whether a student get admission in an university by taking analysis of his marks.

Here the problem is if the student want to join in any university he/she want to wait until he/she get admit card from that college, even if he/she wait we can't give assurance that he may get admission are not.

So, to solve all these problems we are implementing a prediction using Machine Learning.By using these prediction methods we can predict that whether he/she get admission in any University by predicting Chance of getting admission in an University , considering the following details :

GRE Score

TOEFL Score

Statement of Purpose (SOP)

Letter of Recommendation (LOR)

University Ranking

CGPA

Research

By taking analysis of these following values we can predict that we can get admission in any university. These type of prediction can be used for all universities.

***1.4. Industry Profile:***

The Educational Services Industry is composed of establishments that provide instruction and training on a wide variety of subjects.

Theses institutions, including schools, colleges, universities and training centers, are either privately or publicly owned. Private institutions may be further classified as "for-profit" or "not-for-profit". We report on publicly traded, for-profit schools that have a focus on post-secondary education. According to the most recent data provided by the U.S. Department of Education, post-secondary education is being provided to about 18.2 million students. Of that population, some 1.4 million are receiving their education through for-profit schools.

*Industry Dynamics:*

Educational Services is widely considered a counter-cyclical industry. That is to say, typically, when the economy is doing poorly and unemployment is rising, more working adults, as their career prospects start to dim, decide to upgrade their education. This, in turn, leads to higher enrollment and increased profit at the schools. We note that traditional undergraduate education for young students is generally non-cyclical. Culinary arts schools, however, can be labeled as moderately cyclical. Also, certain types of educational institutions do perform largely in sync with the broader economy. For example, providers of information technology instruction benefit in good times, when companies are likely to boost related investment.

There is a growth element to this industry. Education companies are reporting a trend of rising demand from working adults. More and more employers are requiring college degrees for a greater range of jobs. Enrollment rates are tracking higher at most schools. To an 18-year old, thinking about the future, or a 30-year old without a college degree, looking for a career boost, diplomas are becoming the standard rather than the exception.

*Profit Generation:*

Schools are seeing improving enrollment rates in both the traditional and online formats. Traditional schooling caters to young students that have just graduated from high school. Online instruction is more tailored to working adults who prefer to attend classes at evening time or during weekends. Web classes typically carry higher margins than the traditional classroom setting, since professors can connect to a significantly larger number of students and there are no direct brick-and-mortar costs. Globalization also augurs well for the industry. Due to global outsourcing, there is increasing pressure on workers in developed countries to enhance their skills.

There are two other important trends running in the industry's favor. As the U.S. continues to transition from a manufacturing-based economy to one heavily reliant on the service sector, for-profit educators stand to gain from offering courses in information technology, healthcare and business management. And, companies have the opportunity to tap an enormous under-educated segment of the world population. Lucrative markets include China, Brazil and other developing nations. A few educators have already established beachheads in promising overseas regions.

*Regulation and Competition:*

Companies in this industry adhere closely to the Higher Education Act. Compliance with the Act is critical to maintain accreditation; it provides the ability to operate in various states. Accreditation allows a school's students to apply for financial aid under Title IV (low income) of the Act, the Pell Grant, and the G.I. Bill. This is important since a majority of students receive some sort of aid. Notably, the Act, and other regulation, has been fairly successful in shutting down and prohibiting "diploma mills", thus ensuring a fairly high overall quality of instruction.

Competition among these schools for prospective students is intensifying. Expenses of new-student leads and marketing continue to rise. Barriers to entry in this industry are significant. It is very expensive for a potential market entrant to build a school from scratch. The start-up phase can be difficult, especially without a substantial government-supported student base. Also, IT investment can be complex and quite costly, particularly for online operations. Financial constraints can limit a school's ability to expand. Schools prefer to tap the equity market, when their share prices are at elevated levels, rather than issue debt.

*Investment Consideration:*

There are several things to look for when investing in individual companies in the Educational Services Industry. Investors should seek schools that have steadily rising enrollment, which typically leads to strong revenue and earnings growth. Favorable new-student starts and high conversion rates (from inquiries to enrollment) are good indicators of a company's prospects. Schools that have tapped or have plans to enter emerging markets will likely have a "first-mover" advantage and solid long-term growth potential. Institutions that have a sizeable mix of top-quality online business can better lever expanding demand at the bottom line.

**2. Data Collection:**

This dataset is created for prediction of University admissions and the dataset link is below:

<https://www.kaggle.com/mohansacharya/graduate-admissions>

Features in the dataset:

* GRE Scores (290 to 340)
* TOEFL Scores (92 to 120)
* University Rating (1 to 5)
* Statement of Purpose (1 to 5)
* Letter of Recommendation Strength (1 to 5)
* Undergraduate CGPA (6.8 to 9.92)
* Research Experience (0 or 1)
* Chance of Admit (0.34 to 0.97)

Some important information:

* There are 9 columns: Serial No., GRE Score, TOEFL Score, University Rating, SOP, LOR , CGPA, Research, Chance of Admit
* There are no null records.
* There are 500 samples in total.

**3. Methodology:**

***3.1. Exploratory Data Analysis:***  
 This dataset has a binary response (outcome, dependent) variable called admit. There are three predictor variables: GRE Score,TOEFL Score,University rating,SOP,LOR,CGPA and Research . We will treat the variables GRE Score,TOEFL Score ,SOP,LOR and CGPA as continuous. The variable University rating takes on the values 1 through 4. Universities with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.

EDA helps us get a clear idea about our features. It also helps us capture any trend or seasonality of data points. In our case, we can see that “higher the GRE score , TOEFL Score, University rating and CGPA , more the chances of getting admit.”

***3.1.1. Figures and table:***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Serial No. | GRE Score | TOEFL Score | University Rating | SOP | LOR | CGPA | Research | Chance of Admit |
| 1 | 337 | 118 | 4 | 4.5 | 4.5 | 9.65 | 1 | 0.92 |
| 2 | 324 | 107 | 4 | 4 | 4.5 | 8.87 | 1 | 0.76 |
| 3 | 316 | 104 | 3 | 3 | 3.5 | 8 | 1 | 0.72 |
| 4 | 322 | 110 | 3 | 3.5 | 2.5 | 8.67 | 1 | 0.8 |
| 5 | 314 | 103 | 2 | 2 | 3 | 8.21 | 0 | 0.65 |

Table : University Admission Details

***3.2 Statistical techniques and visualization:***

There are lot of classification algorithms out there so how do we know which one is the best?  
We don’t. This is something that comes with experience and expertise. But not to worry, we do have a work around. We can fit multiple algorithms onto our data points and then evaluate the model. This way we can choose the best model which has the least error.

The Statistical techniques that can be used in this model are:

*1.Decision Tree Regression:*

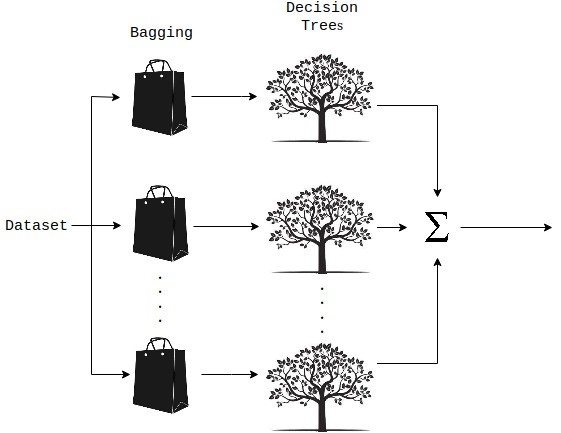
Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node (e.g., Hours Played) represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees can handle both categorical and numerical data.

Libraries that import DecisionTree Regression:

*from sklearn.tree import DecisionTreeRegressor*

*2.Random Forest Regression:*

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called **Bootstrap Aggregation**, commonly knownas**bagging**. What is bagging you may ask? Bagging, in the Random Forest method, involves training each decision tree on a different data sample where sampling is done with replacement.



Libraries that import RandomForest Regression:

*from sklearn.ensemble import RandomForestRegressor*

*3.Linear Regression****:***

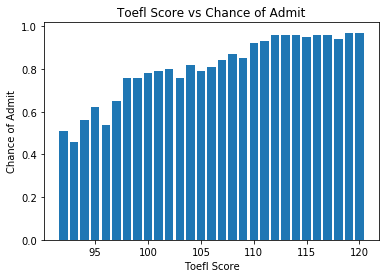
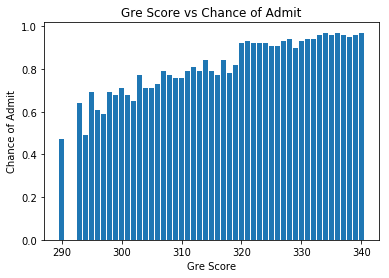
Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. A linear regression line has an equation of the form ***Y = a + bX***, where ***X*** is the explanatory variable and ***Y*** is the dependent variable. The slope of the line is ***b***, and ***a*** is the intercept (the value of ***y***when ***x*** = 0).

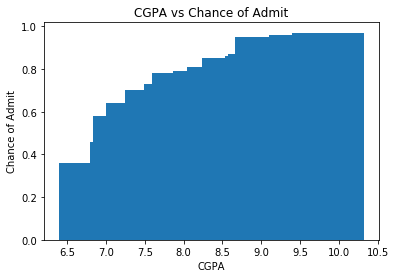
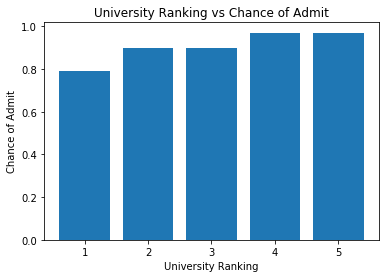
Libraries that import Linear Regression:

*from sklearn.linear\_model import LinearRegression*

***3.3 Data Modeling and visualization:***

Decision trees are graph structures, where each potential decision creates a new node, resulting in a tree-like graph (Quinlan, 1987). In machine learning, decision trees partition the data set in appropriate values until a tree structure has emerged. This process is called recursive partitioning (Strobl, 2009). Decision tree algorithm tries to find the best way to partition the data so that parts are as homogeneous as possible. If a fully homogeneous part is impossible, more common value is chosen.





**4. Findings and Suggestions:**

Here the problem is if the student want to join in any university he/she want to wait until he/she get admit card from that college, even if he/she wait we can't give assurance that he may get admission are not.

So, to solve all these problems we are implementing a prediction using Machine Learning.By using these prediction methods we can predict that whether he/she get admission in any University by predicting Chance of getting admission in an University , considering the following details :

GRE Score

TOEFL Score

Statement of Purpose (SOP)

Letter of Recommendation (LOR)

University Ranking

CGPA

By taking analysis of these following values we can predict that we can get admission in any university. These type of prediction can be used for all universities.

**5. Review of Literature:**

1. <https://edtechdigest.com/2018/05/11/with-machine-learning-predictive-modeling-for-admissions-is-evolving/>

2. <http://sites.tntech.edu/weberle/wp-content/uploads/sites/87/2018/06/NSSR_2013-1.pdf>

3. <https://towardsdatascience.com/predicting-college-acceptance-with-ai-6d8abd702385>

4. <https://towardsdatascience.com/predicting-ms-admission-afbad9c5c599>

5. <http://athena.ecs.csus.edu/~pateljd/images/Admission_prediction_system.pdf>

**6. Conclusion:**

Finally we conclude that instead of waiting to get admit card from university . By using **Machine Learning** prediction we can easly predict wheather a student get admission in a particular university or not.