

EMPLOYABILITY PREDICTOR USING MACHINE LEARNING

-PADMASHRI B V

PROBLEM STATEMENT

In today's job market, students and graduates face lots of struggle to find a job in every industry like in the field of information technology, banking sector, automobile sector and almost all the industries. The problem addressed by the **Employability Predictor** app is the lack of personalized guidance for students who want to improve their employability. Many students graduate with a limited understanding of their strengths and weaknesses in the job market today. The application aims to bridge this gap by providing tailored insights and recommendations based on the user's skills and interest updated in the profile.

MARKET/CUSTOMER/BUSINESS NEED ASSESSMENT

- Today the job market is highly competitive, and students often lack guidance on how to improve their employability
- There is a need of personalized recommendations to make informed decisions about their education and skill development
- The students and recent graduates needs career guidance ,skill improvement,and job market insights
- The businesses can address the gap in the market and provide a valuable service while potentially generating revenue through premium features and partnerships with Top HR's.

TARGET SPECIFICATIONS AND CHARACTERIZATION (Customer Characteristics)

- The primary objective of the Employability Predictor App is to develop machine learning models to calculate personalized employability scores for users
- And create a user-friendly interface for profile creation, employability score prediction, and personalized recommendations
- Our targets are students and recent graduates(age 18-30)from various educational backgrounds
- The app provide insights into job market trends and career pathways
- Ensure data privacy and security in compliance with regulations
- Continuously improve the app through user feedback

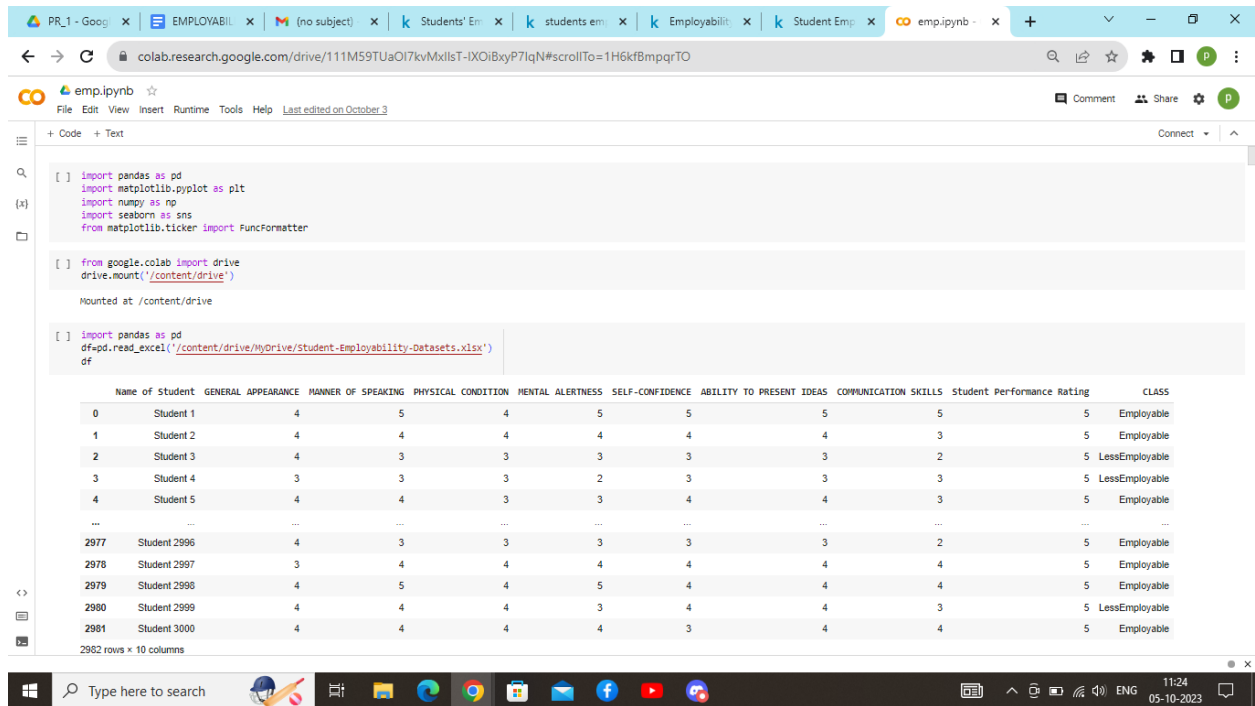
EXTERNAL SEARCH(INFORMATION SOURCES/REFERENCES)

The usage of Student's employability dataset-philippines as base data for this project

Link to this dataset:<https://www.kaggle.com/datasets/anashamoutni/students-employability-dataset>

The dataset is found in kaggle. The dataset consists of Mock job Interview Results of 2982 observations. Dataset was collected from different university agencies in the Philippines. The dataset that was collected is compliant with the Data Privacy Act of the Philippines.

DATASET



```
[ ] import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from matplotlib.ticker import FuncFormatter

[ ] from google.colab import drive
drive.mount('/content/drive')

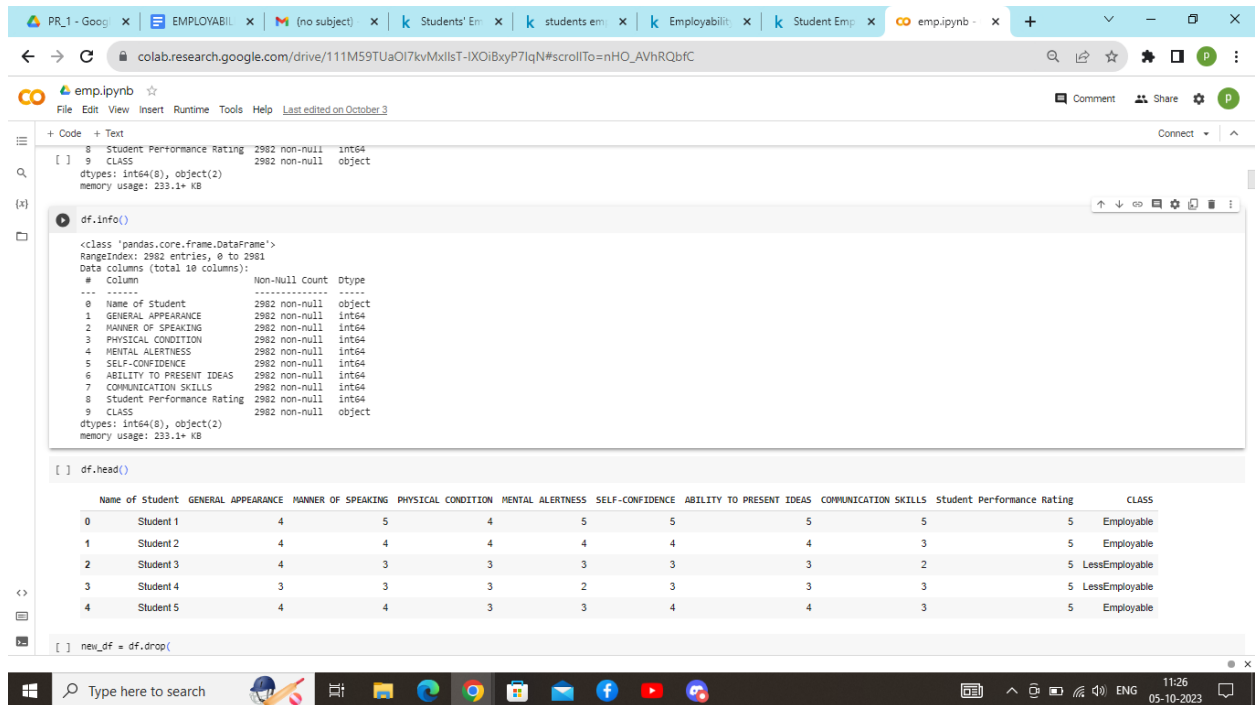
Mounted at /content/drive

[ ] import pandas as pd
df=pd.read_excel('/content/drive/myDrive/Student-Employability-Datasets.xlsx')
df
```

	Name of Student	GENERAL APPEARANCE	MANNER OF SPEAKING	PHYSICAL CONDITION	MENTAL ALERTNESS	SELF-CONFIDENCE	ABILITY TO PRESENT IDEAS	COMMUNICATION SKILLS	Student Performance Rating	CLASS
0	Student 1	4	5	4	5	5	5	5	5	Employable
1	Student 2	4	4	4	4	4	4	3	5	Employable
2	Student 3	4	3	3	3	3	3	2	5	LessEmployable
3	Student 4	3	3	3	2	3	3	3	5	LessEmployable
4	Student 5	4	4	3	3	4	4	3	5	Employable
...
2977	Student 2996	4	3	3	3	3	3	2	5	Employable
2978	Student 2997	3	4	4	4	4	4	4	5	Employable
2979	Student 2998	4	5	5	5	4	4	4	5	Employable
2980	Student 2999	4	4	4	3	4	4	3	5	LessEmployable
2981	Student 3000	4	4	4	4	3	4	4	5	Employable

2982 rows x 10 columns

INFORMATION OF THE DATASET



```
[ ] df.info()

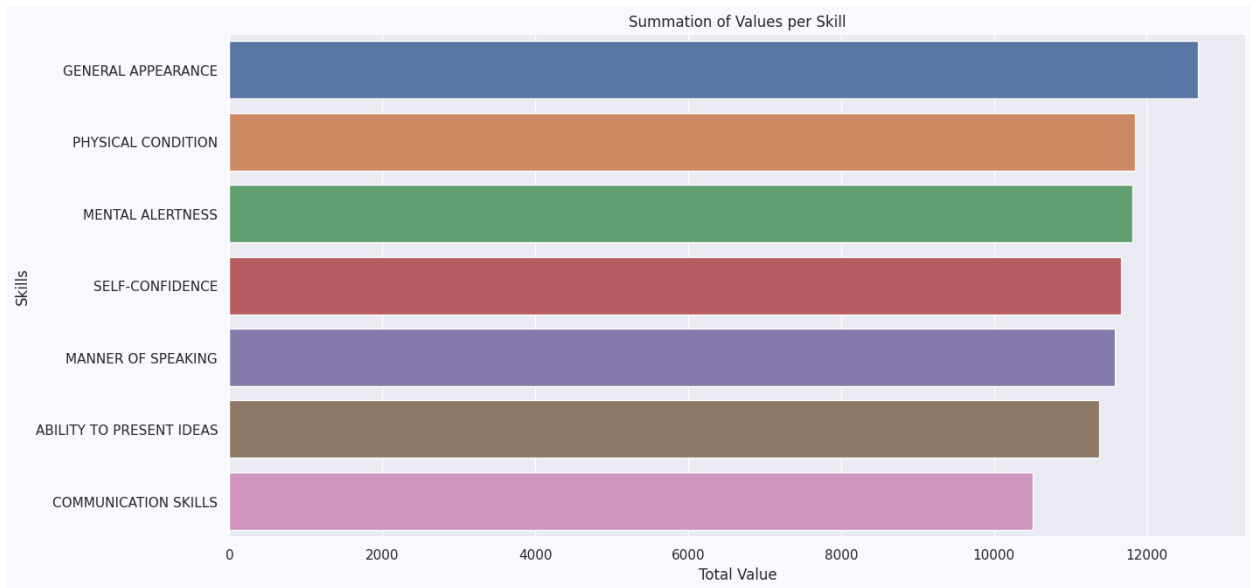
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2982 entries, 0 to 2981
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Name of Student       2982 non-null   object
 1   GENERAL APPEARANCE    2982 non-null   int64
 2   MANNER OF SPEAKING    2982 non-null   int64
 3   PHYSICAL CONDITION    2982 non-null   int64
 4   MENTAL ALERTNESS     2982 non-null   int64
 5   SELF-CONFIDENCE      2982 non-null   int64
 6   ABILITY TO PRESENT IDEAS 2982 non-null   int64
 7   COMMUNICATION SKILLS  2982 non-null   int64
 8   Student Performance Rating 2982 non-null   int64
 9   CLASS                2982 non-null   object
dtypes: int64(8), object(2)
memory usage: 233.1+ KB

[ ] df.head()

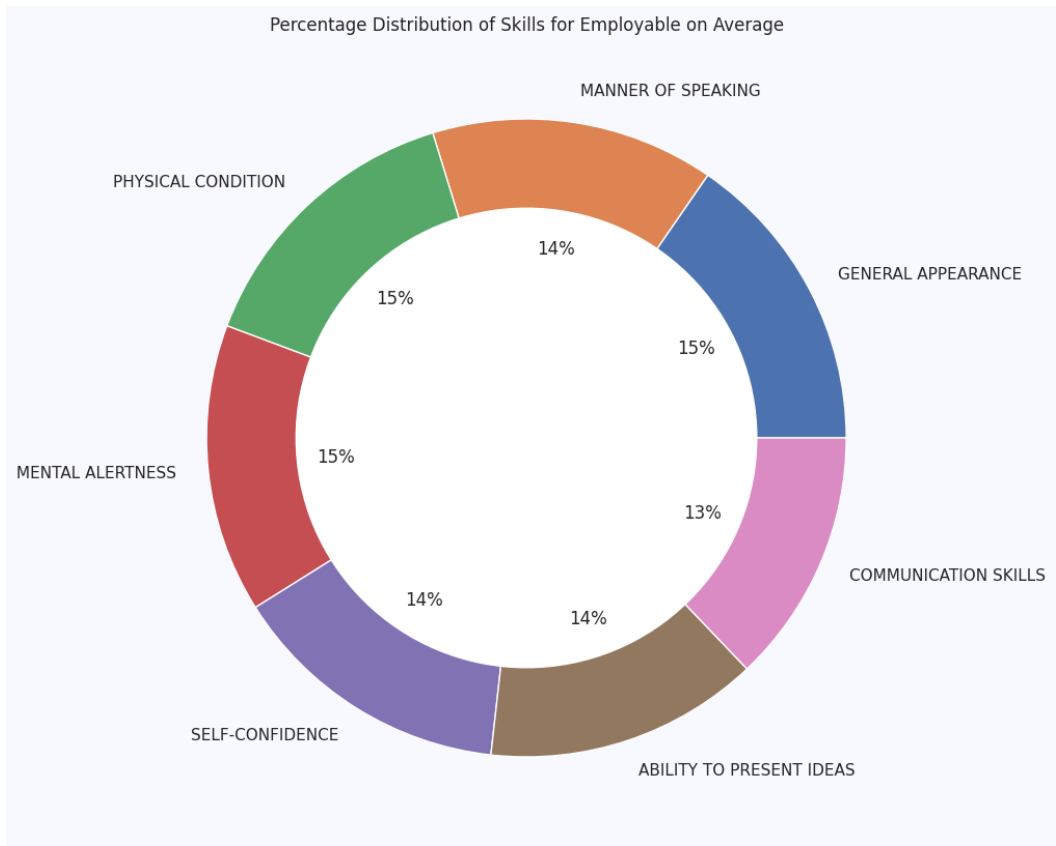
Name of Student  GENERAL APPEARANCE  MANNER OF SPEAKING  PHYSICAL CONDITION  MENTAL ALERTNESS  SELF-CONFIDENCE  ABILITY TO PRESENT IDEAS  COMMUNICATION SKILLS  Student Performance Rating  CLASS
0      Student 1                4                5                4                5                5                5                5                5      Employable
1      Student 2                4                4                4                4                4                4                3                5      Employable
2      Student 3                4                3                3                3                3                3                2                5      LessEmployable
3      Student 4                3                3                3                2                3                3                3                5      LessEmployable
4      Student 5                4                4                3                3                4                4                3                5      Employable

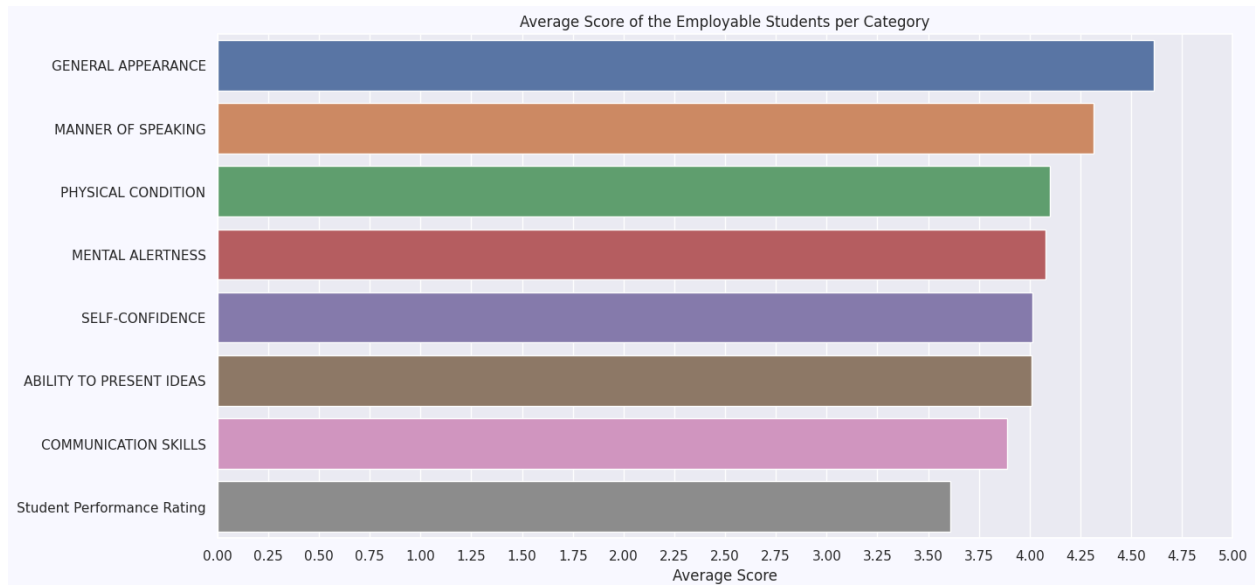
[ ] new_df = df.drop()
```

BENCHMARKING

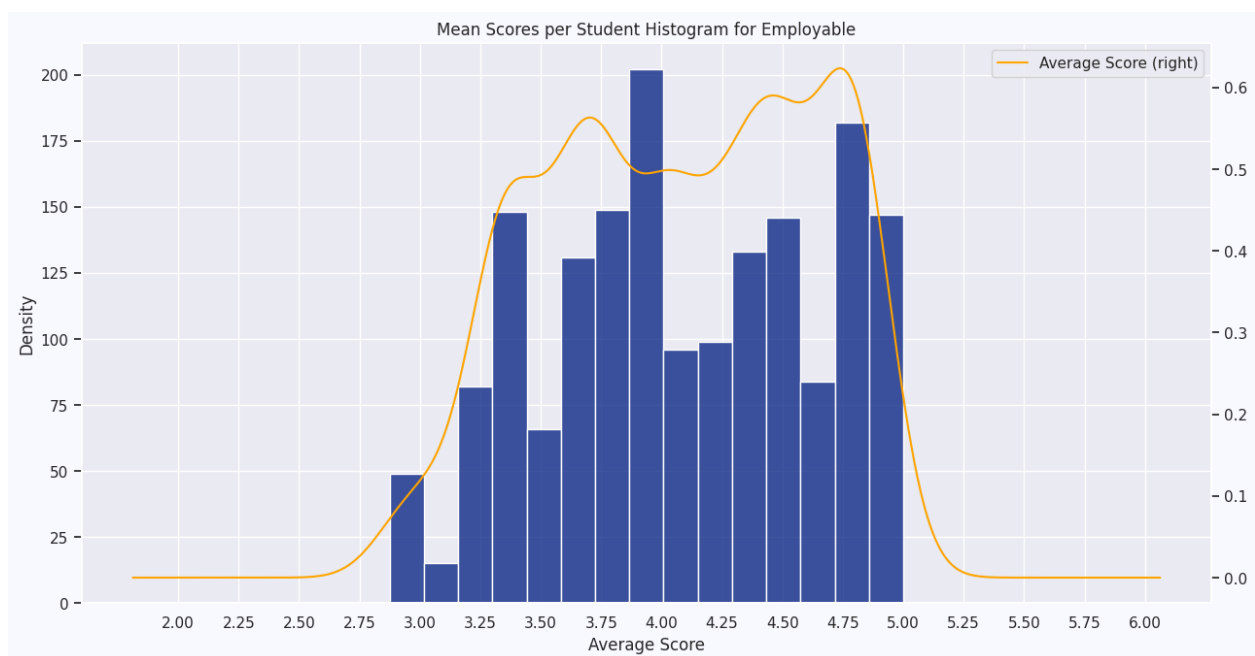


students are lacking in skills but best in general appearance

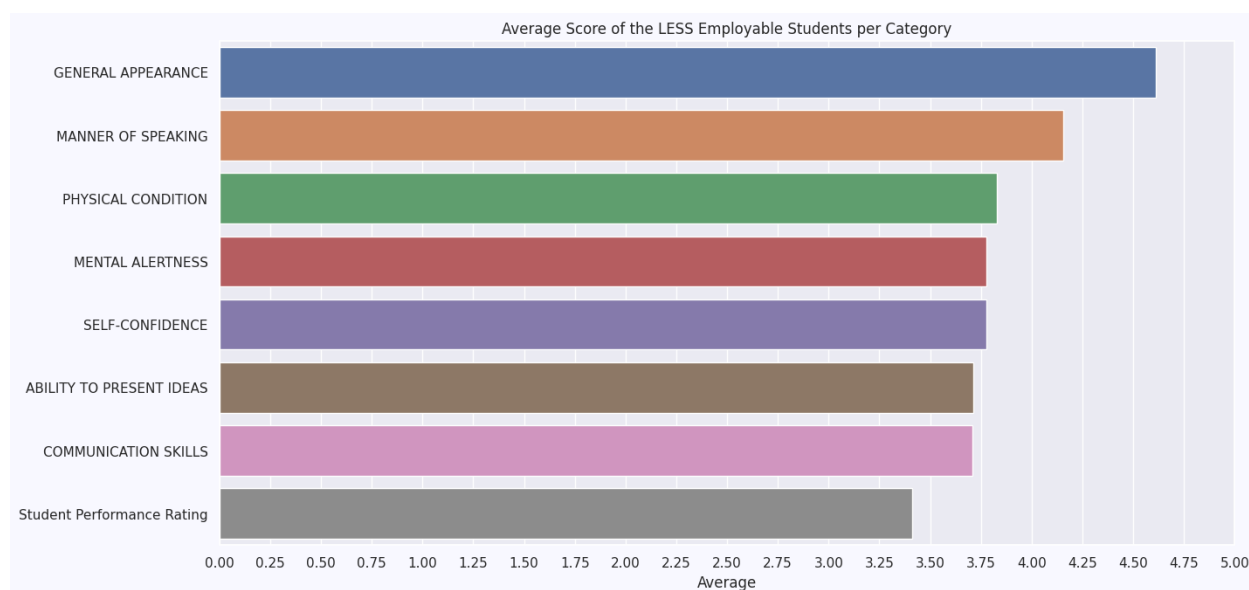




We can see that the mean scores for employable students are capable of reaching a mark of 4 and above for most of the skills

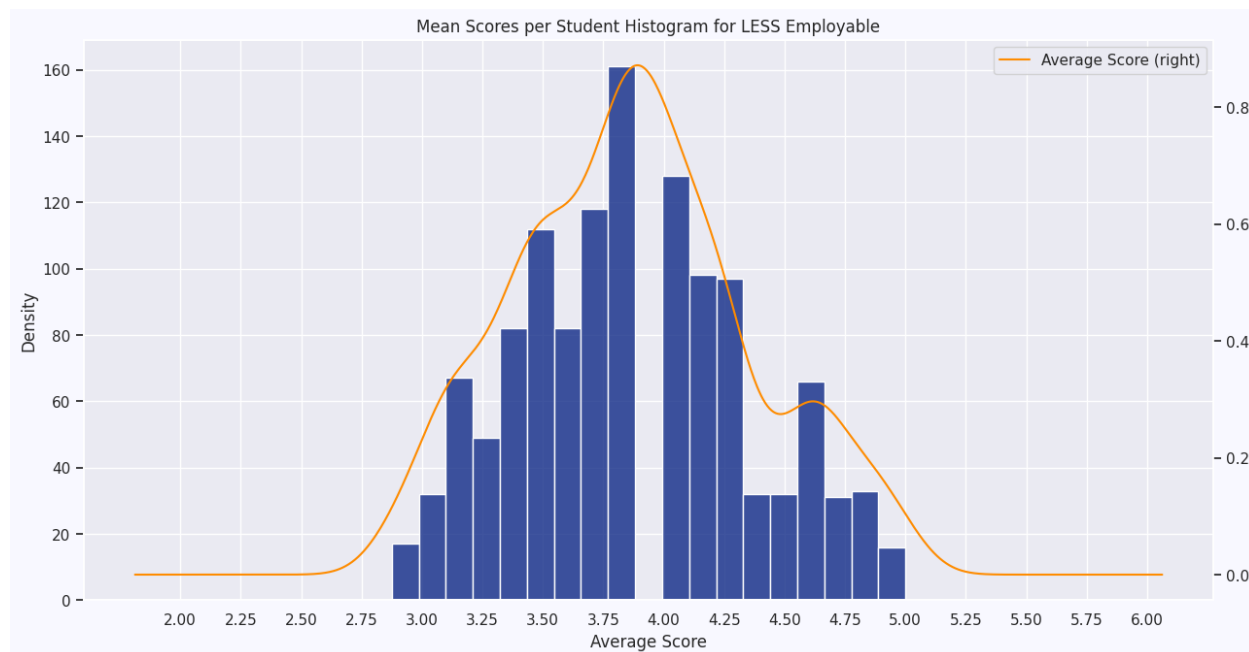


The Histogram shows that the minimum scores for the Employable can be at least 3.0

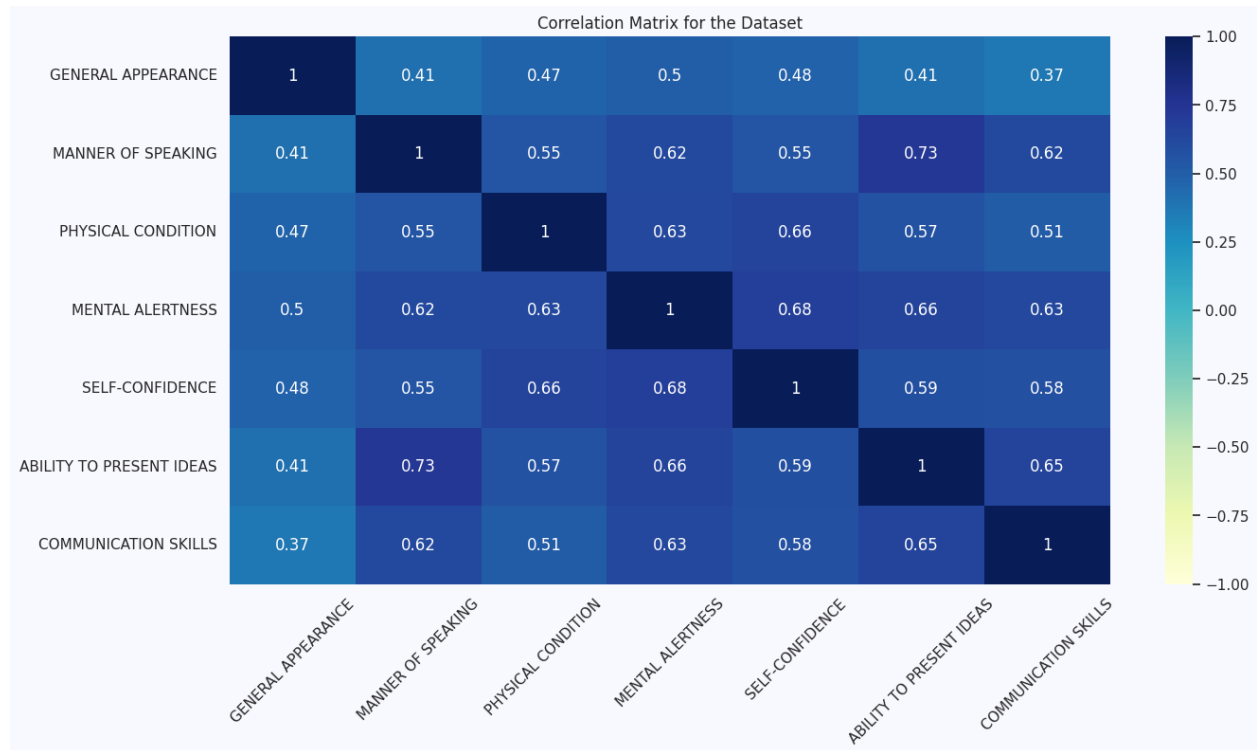


Comparing this to the Employable most of the Categories in Less Employable failed to reach an average score of 4, except for general appearance. Whereas in Employable, 5 categories were able to reach an average score of 4.

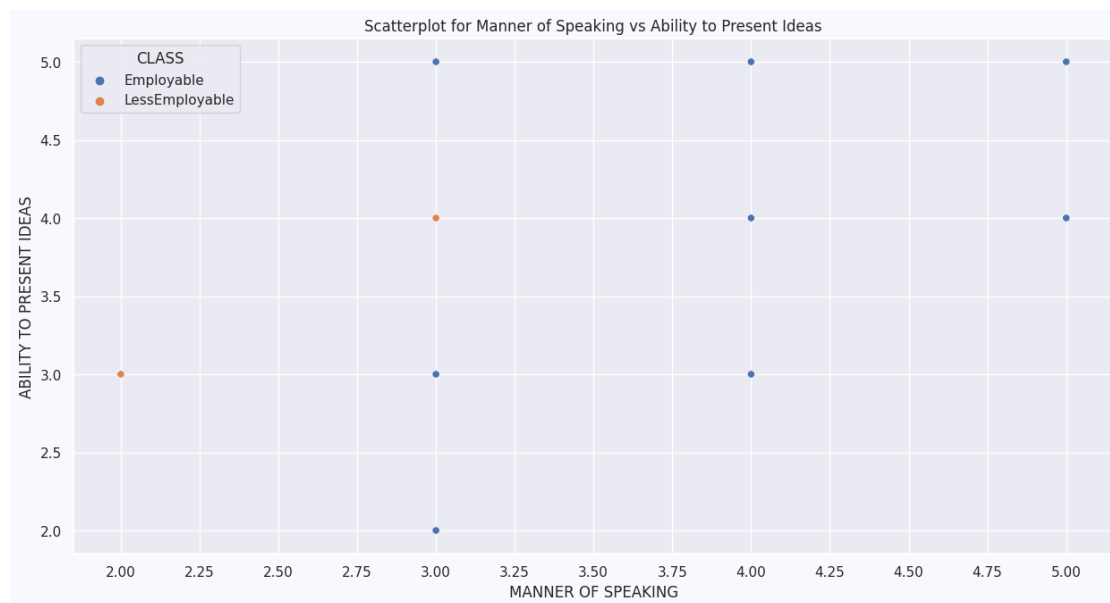
In both charts we can see that Students are lacking of Communication skills, but is excelling at General Appearance

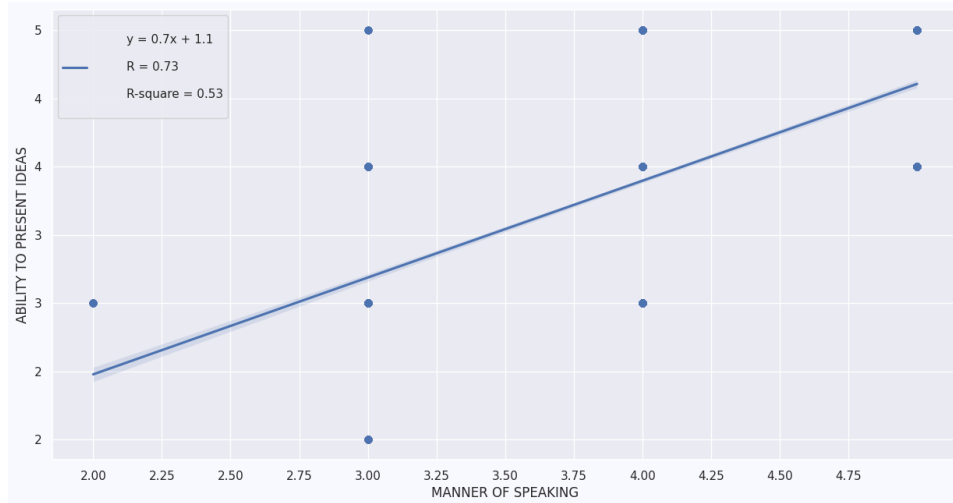


Surprisingly, Less Employable Students can also reach an average score of 4 and higher. This shows that it is possible to be Less Employable even if you can reach an average score higher than 4.



We can see that there is a strong correlation between Manner of Speaking and The Ability to present ideas with an R of 0.73 and R-square of 0.53





APPLICABLE PATENTS

- **Conduct a patent search to ensure there are no infringements in the proposed machine learning models or algorithms**
- **Explore licensing options for any required third-party technologies**

APPLICABLE REGULATIONS

- **Ensuring compliance with data privacy regulations like GDPR,CCPA while handling the data**
- **Adhere to guidelines for personalized recommendations**

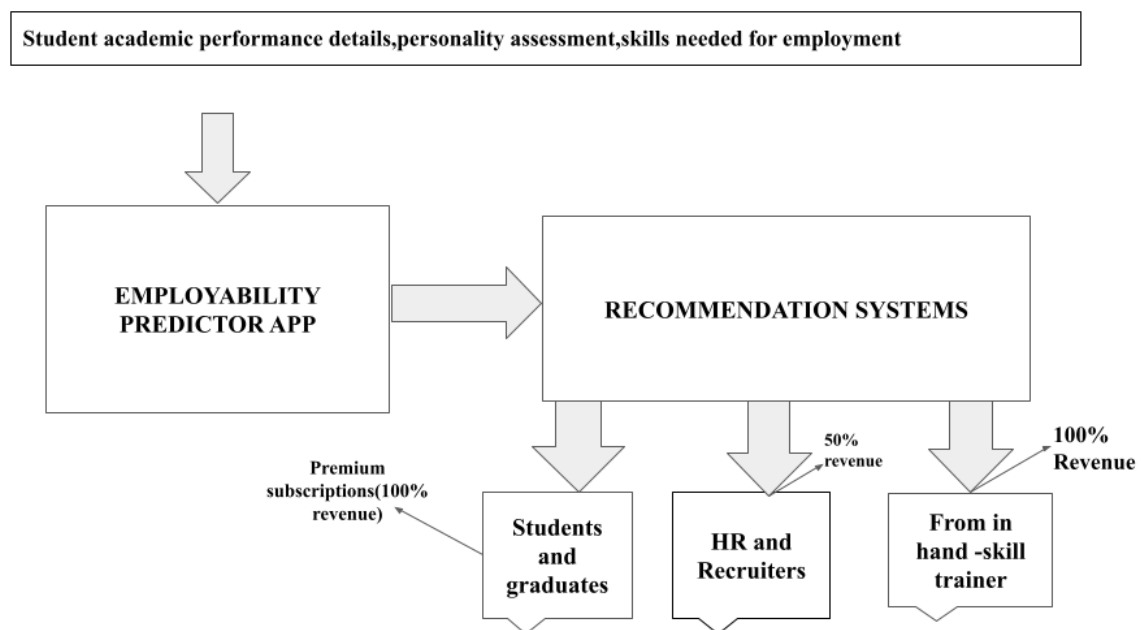
APPLICABLE CONSTRAINTS

- **Limited budget for app development and marketing**
- **Need for a multidisciplinary team with expertise in machine learning , App development**

CONCEPT GENERATION

- The idea comes from the need for personalized career guidance in a competitive job market. Inspired by the success of machine learning -driven recommendation systems in various domains. The product requires building a machine learning model from scratch.

FINAL PRODUCT PROTOTYPE



The product takes the following functions to perfect and provide a good result.

Back-end

Model Development: This must be done before releasing the service. A lot of manual supervised machine learning must be performed to optimize the automated tasks.

1. Performing EDA to realize the dependent and independent features.
2. Algorithm training and optimization must be done to minimize overfitting of the model and hyperparameter tuning.

Front End

- 1. Different user interface:** The user must be given many options to choose from in terms of parameters. This can only be optimized after a lot of testing and analysis of all the edge cases.
- 2. Interactive visualization of the data extracted from the trained models** will return raw and inscrutable data. This must be present in an aesthetic and an “easy to read” style.
- 3. Feedback system:** A valuable feedback system must be developed to understand the customer’s needs that have not been met. This will help us train the models constantly.

PROTOTYPE SELECTION

A)FEASIBILITY

This project can be developed and deployed within a few years as SaaS(Software as a service) to use

B)VIABILITY

In today’s world, unemployment is one of the major issues in our country . So this kind of development in these industries can bring many positive approaches towards this issue .

C)MONETIZATION

This can be monetized by Basic subscription model and Premium model

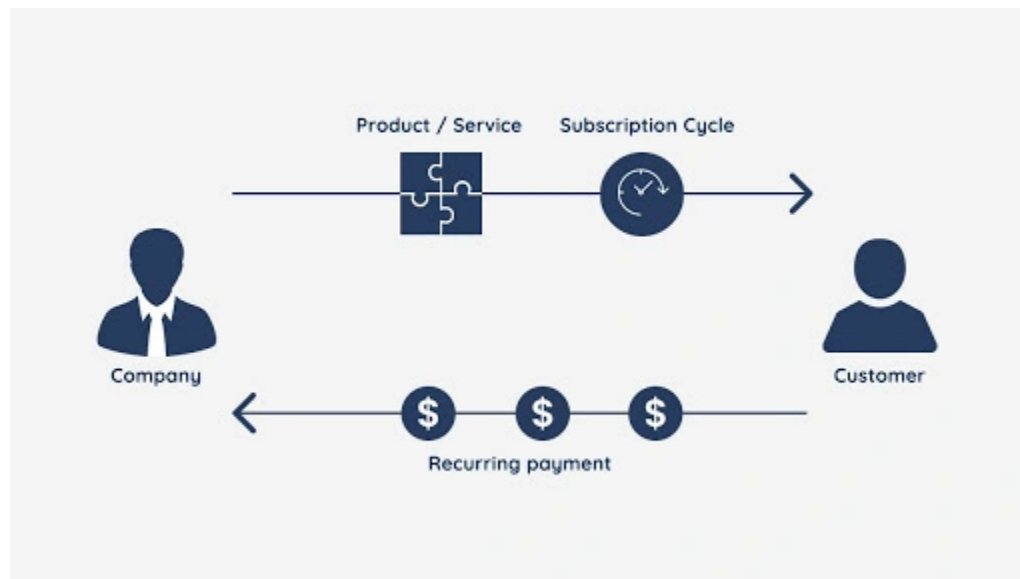
PROTOTYPE DEVELOPMENT

The “Employability predictor app” will assess student’s employability using machine learning models that consider academic performance ,skills, and extracurricular activities to provide personalized employability scores and recommendations for skill development ,career pathways, and job market insights.The concept can be developed by using The appropriate API (flask in this case) and using Django as framework for the same and for its deployment , The cloud services has to be chosen accordingly to the need.

CODE

IMPLEMENTATION:<https://colab.research.google.com/drive/1V3EKtM4n2g8qhBt51umv3gVRfu23YpMn?usp=sharing>

BUSINESS MODEL



- **FREE ASSESSMENT:** offering basic employability assessment and recommendations for free for two weeks and charge for premium features like in-depth career counseling and personalized coaching
- **BASIC SUBSCRIPTION MODEL:** charging a monthly or annual subscription fee for access to advanced features with fare charges compared to others
- **PREMIUM MODEL:** Direct collaboration with HR and recruiters and get trained by in hand practice by No.1 skill trainers to make available personalized jobs and personalized skill development for clients

FINANCIAL EQUATION



According to the above image , we can see the global recruitment software market size from 2015 to 2025 , we can see massive growth of the software market in recruitment .

So if we follow the above trend to our problem, it would be advisable to price our service around Rs.2000 for one month premium subscription and Rs.999 for one month subscription for basic model subscription, we also offer free trial for two weeks of basic model, if clients are satisfied they can opt for basic or premium subscription. Once the client base increases we can either increase the price, according to month.

Let's assume that the duration of developing the ML model takes about 1 to 3 weeks and the cost for producing the model is the salary of the members of the team. Let there be three ML engineers and two full stack web developers. Let the salary of the ML engineers be 'ml' and the full stack web developers be 'fs'. Let's keep the price of our product m as 6000, So the total cost $c = 3*ml + 2fs$. So the profit or financial equation will look like this

$$y = m \cdot x(t) - c$$

$x(t)$ represents the total customer base growth on time,

y represents profit

Let assume the salary and production cost of c as 100000

$$y = 6000 \cdot x(t) - 100000$$

CONCLUSION

The "Employability Predictor" app aims to address the critical need for personalized career guidance in today's competitive job market. By leveraging machine learning, it provides actionable insights and recommendations to students and job seekers, empowering them to make informed decisions about their education and career paths. The app's potential for monetization and the growing demand for such services make it a promising venture in the education and career counseling sector.