# IDENTIFYING PATTERNS AND TRENDS IN CAMPUS PLACEMENT DATA USING MACHINE LEARNING

# 1.INTRODUCTION

### 1.1 **OVERVIEW**:

In Placement Prediction system predicts the probability of an undergraduate students getting placed in a company by applying classification algorithms such as Decision tree and Random forest. The main objective of this model is to predict whether the student he/she gets placed or not in campus recruitment. For this the data consider is the academic history of student like overall percentage, backlogs, credits. The algorithms are applied on the previous years data of the students.

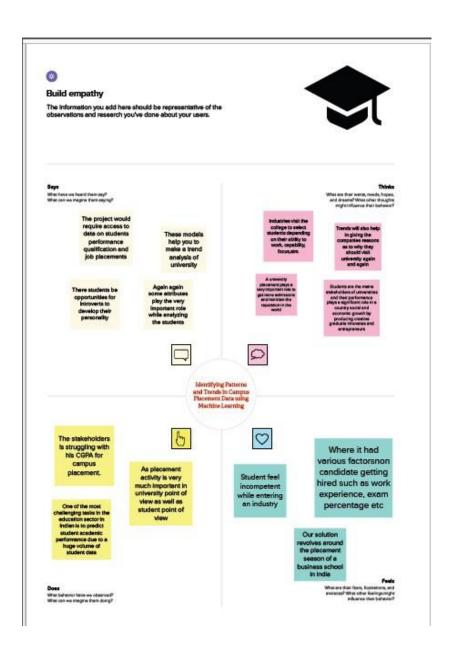
#### 1.2 Purpose:

The objective is to predict the students getting placed for the current year by analyzing the data collected from previous year's students. This model is proposed with an algorithm to predict the same. These models help you to make a trend analysis of university placementsdata, to predict a placement rate for the students of an upcoming year which will help the university to analyze the performance during placements.

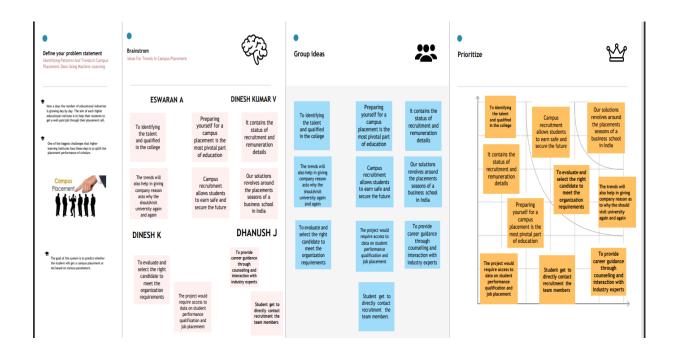
This proposed model is also compared with other traditional classification algorithms such as Decision tree and Random forest with respect to accuracy, precision and recall. From the results obtained it is found that the proposed algorithm performs significantly better in comparison with the other algorithms mentioned.

# 2. PROBLEM DEFINITION AND DESIGN THINKING

### 2.1 Empathy Map:



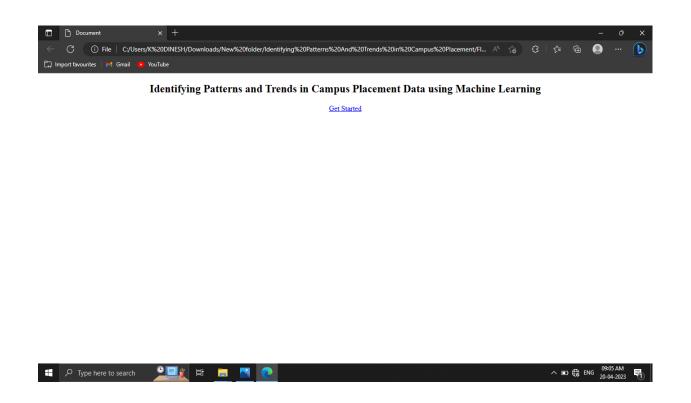
### 2.2 Ideation & Brainstorming Map:



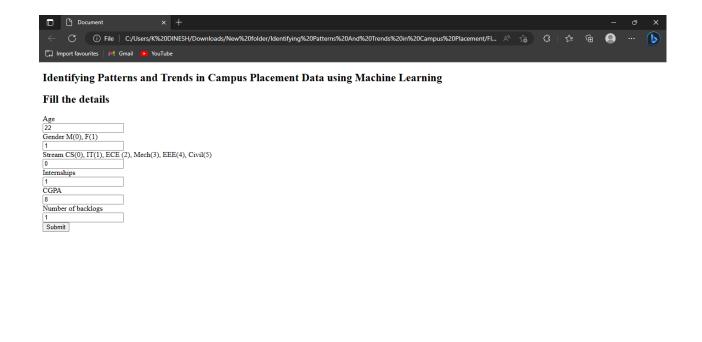
# 3. RESULT

The algorithms of machine learning we have discussed are can used to find the trend of placement, which will be helpful for university to get more admission in future.

### 3.1 Home Page/index1:

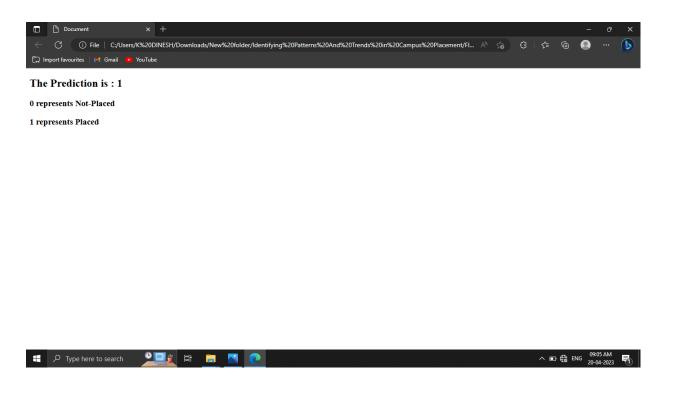


# 3.2 Fill the Details/index:



^ ■ € ENG 20.04.2022

# 3.3 Predicted page (Final output)/second page:



# 4. TRAILHEAD PROFILE LINK

Team Lead- <a href="https://trailblazer.me/id/eswaran77">https://trailblazer.me/id/eswaran77</a>

Team Member 1- <a href="https://trailblazer.me/id/dineshkumarv18">https://trailblazer.me/id/dineshkumarv18</a>

Team Member 2- <a href="https://trailblazer.me/id/dineshkd07">https://trailblazer.me/id/dineshkd07</a>

Team Member 3- <a href="https://trailblazer.me/id/ddhanush23">https://trailblazer.me/id/ddhanush23</a>

# 5. ADVANTAGES AND DISADVANTAGES

#### **Advantages:**

- ✓ By analyzing campus placement data, recruiters can identify which
  universities or program produces the most successful job candidates. This
  can help recruiters to focus their efforts on these institutions and programs,
  and to tailor their recruitment strategies accordingly.
- ✓ By examining the skills and qualifications of successful job candidates, employers can identify areas where their current workforce may be lacking. This can help employers to develop targeted training and development programs to address these skill gaps.

- ✓ Campus placement data can also be used to monitor diversity and inclusion efforts.
- ✓ Examining campus placement data over time can provide insights into industry trends and changes in the job market.
- ✓ By analyzing the success rates of graduates from different educational programs, employers can evaluate the effectiveness of those programs in preparing students for the job market.

#### **Disadvantages:**

- ✓ Campus placement may only represent a small sample of the overall job market.
- ✓ Campus placement data may be influenced by various biases, such as the preferences of recruiters or the characteristics of the universities or program being studied.
- ✓ Campus placement data may be incomplete or inaccurate due to factors such as incomplete reporting or data entry errors.
- ✓ Campus placement data may not provide enough context to fully understand the factors that contribute to job placement success.

✓ The collection and analysis of campus placement data may raise ethical concerns related to privacy and confidentially. Organizations must ensure that they are collecting and analyzing the data in a responsible and ethical manner.

# 6. APPLICATIONS

- ✓ Campus placement data can help organizations to develop more effective recruitment strategies.
- ✓ BY identifying which universities, programs, or majors produce the most successful job candidates, organizations can focus their recruitment efforts on these areas.
- ✓ Campus placement data can be used to identify skills gaps in the workforce.
- ✓ Campus placement data can be used to monitor diversity and inclusion efforts.
- ✓ Campus placement data can provide insights into industry trends and changes in the job market.
- ✓ Campus placement data can be used to identify potential future leaders within the organizations.

# 7. CONCLUSION

The campus placement activity is incredibly a lot of vital as institution point of view as well as student point of view. In this regard to improve the student's performance, a work has been analyzed and predicted using the classification algorithms Decision Tree and the Random forest algorithm to validate the approaches. The algorithms are applied on the data set and attributes used to build the model. The accuracy obtained after analysis for Decision tree is 84% and for the Random Forest is 86%. Hence, from the above said analysis and prediction it's better if the Random Forest algorithm is used to predict the placement results.

# 8. FUTURE SCOPE

Moreover from the study, the researcher concludes that most of the institutions campus placement is not taken seriously. The awareness of such campus hiring should be taught to the students. The students also should consider the importance of campus placement as a part of academics and should have the ability to create academic balance. The significance of Off-campus drives should be taught to the students to crack the campus efficiently. The students should also develop the habit of reading books and journals to prepare themselves for campus placement. The organization should find ways to develop their process of recruitment which in turn helps in an increase in return on investment of campus hiring.

# 9. APPENDIX

#### **Source Code:**

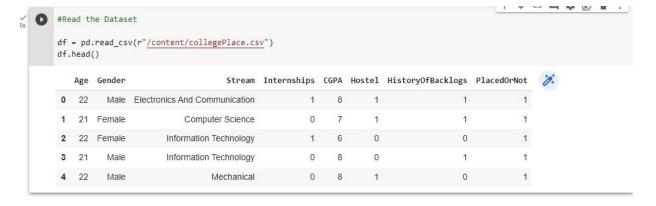
# **Milestone 1:**

**Data Collection & Preparation** 

#### Importing the libraries:

```
_{0s}^{\checkmark} [1] #TASK 2 [DATA COLLECTION & PREPARATION]
                                                                                                        ↑ ↓ ⊖ 目 ‡ 见 🔋 :
   #importing the libraries
        {\tt import\ numpy\ as\ np}
        import pandas as pd
        import os
        import seaborn as sns
        import matplotlib.pyplot as plt
        import warnings
        warnings.filterwarnings('ignore')
        from sklearn import svm
        from sklearn.svm import SVC
        from sklearn.metrics import accuracy_score
        from sklearn.neighbors import KNeighborsClassifier
        from sklearn import metrics
        from sklearn.model_selection import cross_val_score
        from sklearn import preprocessing
        from sklearn.model_selection import train_test_split
        from sklearn.preprocessing import StandardScaler
        import joblib
        from sklearn.metrics import accuracy_score
```

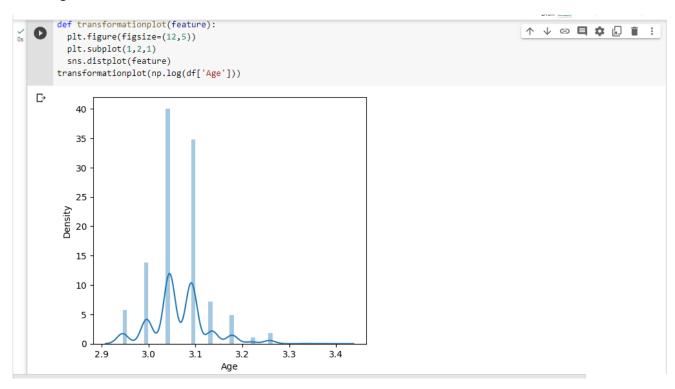
#### Read the Dataset:



#### Handling missing values:



#### Handling outliers:



#### Handling Categorical Values:

```
# Handling Categorical Values

df = df.replace(['Male'],[0])
df = df.replace(['Female'],[1])

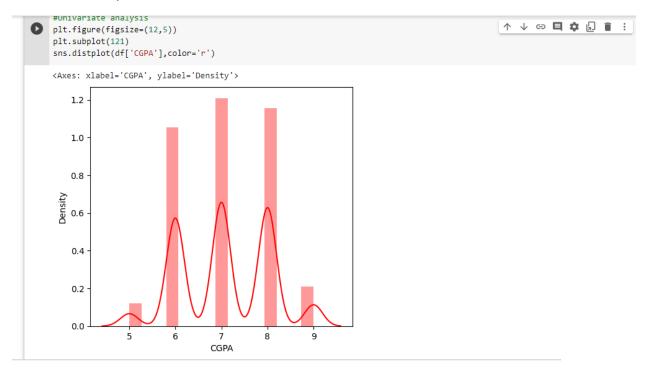
df = df.replace(['Computer Science','Information Technology','Electronics And Communication','Mechanical','Electrical','C:
```

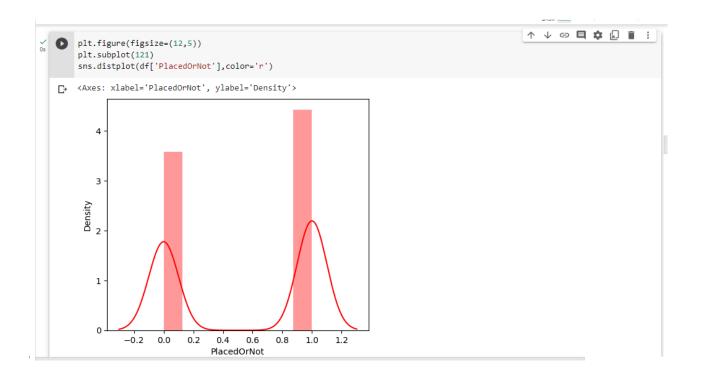
D	df = d	df.dro	p(['Hos	tel'], a	xis=1)					<b></b>	<b>↓</b> G	· <b>=</b>	₽	<u>F</u> ]	ĺ
	df														
		Age	Gender	Stream	Internships	CGPA	HistoryOfBacklogs	PlacedOrNot	<b>*</b>						
	0	22	0	2	1	8	1	1							
	1	21	1	0	0	7	1	1							
	2	22	1	1	1	6	0	1							
	3	21	0	1	0	8	1	1							
	4	22	0	3	0	8	0	1							
					•••										
	2961	23	0	1	0	7	0	0							
	2962	23	0	3	1	7	0	0							
	2963	22	0	1	1	7	0	0							
	2964	22	0	0	1	7	0	0							
	2965	23	0	5	0	8	0	1							
	2966 rd	ows x	7 columns												

# Milestone 2:

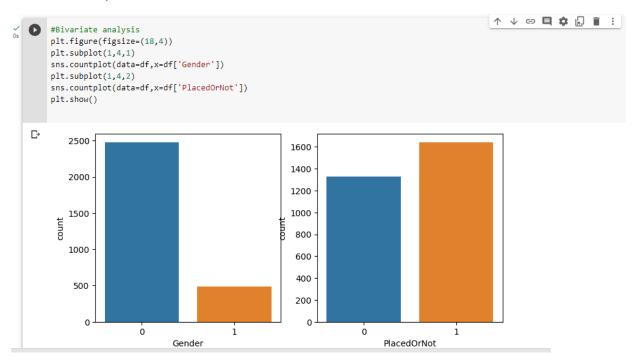
# **Exploratory Data Analysis**

#### Uni-variate analysis:

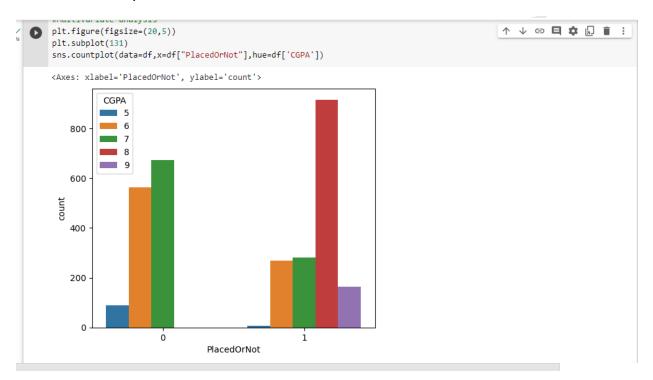


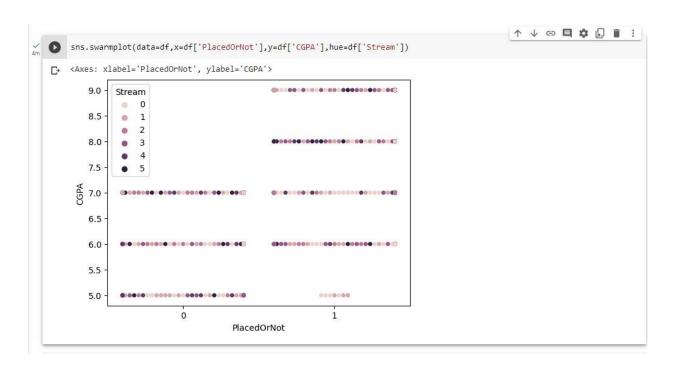


#### Bi-variate analysis:



#### Multi-variate analysis:





#### Scaling the data:

```
v    [16] #Scaling the data

sc=StandardScaler()
x_bal=sc.fit_transform(df)
x_bal=pd.DataFrame(x_bal)
```

#### Splitting the data into train and test:

```
[17] #Splitting the data into train and test

x=x_bal
y=df['PlacedOrNot']

X_train, X_test, Y_train, Y_test = train_test_split(x,y, test_size = 0.2, stratify=y, random_state=2)
```

# Milestone 3: Model Building

#### SVM model:

#### KNN model:

```
best_k = {"Regular":0}
best_score = {"Regular":0}
for k in range(3,50,2):
    knn_temp = KNeighborsClassifier(n_neighbors=k)
    knn_temp.fit(X_train, Y_train)
    knn_temp_pred = knn_temp.predict(X_test)
    score = metrics.accuracy_score(Y_test, knn_temp_pred) * 100
if score >= best_score["Regular"] and score < 100:
    best_score["Regular"] = score
    best_k["Regular"] = k
else:
    best_score["Regular"] = score
    best_k["Regular"] = k
break</pre>
```

```
print("---Results---\nK: {}\nScore: {}".format(best_k, best_score))
knn = KNeighborsClassifier(n_neighbors=best_k["Regular"])
knn.fit(X_train, Y_train)
knn_pred = knn.predict(X_test)
testd = accuracy_score(knn_pred,Y_test)
print(testd)

---Results---
K: {'Regular': 7}
Score: {'Regular': 87.20538720538721}
0.8720538720538721
```

#### Artificial neural network model:

```
#Artificial neural network model

import tensorflow as tf

from tensorflow import keras

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

[23] classifier = Sequential()

classifier.add(keras.layers.Dense(6,activation = 'relu',input_dim=6))

classifier.add(keras.layers.Dropout(0.50))

classifier.add(keras.layers.Dense(6,activation = 'relu'))

classifier.add(keras.layers.Dropout(0.50))

classifier.add(keras.layers.Dense(1,activation = 'sigmoid'))

| loss_1 = tf.keras.losses.BinaryCrossentropy()

classifier.compile(optimizer = 'Adam', loss=loss_1, metrics=['accuracy'])
```

```
↑ ↓ © 目 $ 见 i :
classifier.fit(X_train, Y_train, batch_size=20, epochs=100)
    Epoch 1/100
    Epoch 2/100
    119/119 [================= ] - 0s 2ms/step - loss: 0.7836 - accuracy: 0.4798
    Epoch 3/100
    Epoch 4/100
    119/119 [================== - 0s 2ms/step - loss: 0.7029 - accuracy: 0.4751
    Epoch 5/100
    119/119 [================== ] - 0s 2ms/step - loss: 0.6952 - accuracy: 0.4857
    Epoch 6/100
    Epoch 7/100
    119/119 [================== ] - 0s 2ms/step - loss: 0.6936 - accuracy: 0.5405
    Epoch 8/100
    Epoch 9/100
    119/119 [============ ] - 0s 2ms/step - loss: 0.6905 - accuracy: 0.5527
    Epoch 10/100
    Epoch 11/100
  Epoch 12/100
  Epoch 13/100
  119/119 [================== ] - 0s 2ms/step - loss: 0.6873 - accuracy: 0.5527
  Epoch 14/100
  Epoch 15/100
  119/119 [================== ] - 0s 2ms/step - loss: 0.6887 - accuracy: 0.5527
  Epoch 16/100
  119/119 [-----] - 0s 2ms/step - loss: 0.6875 - accuracy: 0.5527
  Epoch 17/100
  Epoch 18/100
  119/119 [================= ] - 0s 2ms/step - loss: 0.6886 - accuracy: 0.5527
  Epoch 19/100
  Epoch 20/100
  Epoch 21/100
   119/119 [============== ] - 0s 2ms/step - loss: 0.6882 - accuracy: 0.5527
   Epoch 22/100
   Epoch 23/100
   119/119 [=============== ] - 0s 2ms/step - loss: 0.6881 - accuracy: 0.5527
   Epoch 24/100
   119/119 [================ ] - 0s 2ms/step - loss: 0.6878 - accuracy: 0.5527
   Epoch 25/100
   119/119 [============== ] - 0s 2ms/step - loss: 0.6880 - accuracy: 0.5527
   Epoch 26/100
   119/119 [============= ] - 0s 2ms/step - loss: 0.6875 - accuracy: 0.5527
   Epoch 27/100
   119/119 [================ ] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
   Epoch 28/100
   119/119 [============== ] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
   Epoch 29/100
   119/119 [================= ] - 0s 2ms/step - loss: 0.6878 - accuracy: 0.5527
   Epoch 30/100
   119/119 [=============== ] - 0s 2ms/step - loss: 0.6878 - accuracy: 0.5527
```

```
Epoch 31/100
119/119 [============ ] - 0s 2ms/step - loss: 0.6876 - accuracy: 0.5527
Epoch 32/100
119/119 [============ ] - 0s 2ms/step - loss: 0.6876 - accuracy: 0.5527
Epoch 33/100
119/119 [================= ] - 0s 2ms/step - loss: 0.6878 - accuracy: 0.5527
Epoch 34/100
Epoch 35/100
119/119 [============ ] - 0s 2ms/step - loss: 0.6876 - accuracy: 0.5527
Epoch 36/100
119/119 [================ ] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
Epoch 37/100
Epoch 38/100
Epoch 39/100
Epoch 40/100
119/119 [=========== - 0s 2ms/step - loss: 0.6875 - accuracy: 0.5527
Epoch 41/100
Epoch 42/100
Epoch 43/100
119/119 [============ ] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
Epoch 44/100
119/119 [=============== ] - 0s 2ms/step - loss: 0.6875 - accuracy: 0.5527
Epoch 45/100
Epoch 46/100
Epoch 47/100
Epoch 48/100
Epoch 49/100
Epoch 50/100
```

```
Epoch 51/100
Epoch 52/100
  119/119 [====
       Epoch 53/100
  Epoch 54/100
  119/119 [========== ] - 0s 2ms/step - loss: 0.6873 - accuracy: 0.5527
  Epoch 55/100
  Epoch 56/100
  119/119 [=========== ] - 0s 2ms/step - loss: 0.6872 - accuracy: 0.5527
  Epoch 57/100
  119/119 [========== ] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
  Epoch 58/100
  119/119 [=========== ] - 0s 2ms/step - loss: 0.6873 - accuracy: 0.5527
  Epoch 59/100
 Epoch 60/100
  119/119 [============] - 0s 2ms/step - loss: 0.6879 - accuracy: 0.5527
  Epoch 61/100
  Epoch 62/100
  119/119 [============ ] - 0s 2ms/step - loss: 0.6874 - accuracy: 0.5527
  Epoch 63/100
  Epoch 64/100
_____119/119 [=======================] - 0s 2ms/step - loss: 0.6877 - accuracy: 0.5527
Epoch 65/100
Epoch 66/100
Epoch 67/100
Epoch 68/100
Epoch 69/100
Epoch 70/100
Epoch 71/100
Epoch 72/100
Epoch 73/100
Epoch 74/100
```

Epoch 75/100

```
Epoch 76/100
 Epoch 77/100
 Epoch 78/100
 Epoch 79/100
 119/119 [-----] - 0s 2ms/step - loss: 0.6828 - accuracy: 0.5527
 Epoch 80/100
 119/119 [============ ] - 0s 2ms/step - loss: 0.6842 - accuracy: 0.5527
 Epoch 81/100
 119/119 [============ ] - 0s 2ms/step - loss: 0.6841 - accuracy: 0.5527
 Epoch 82/100
 119/119 [============= - 0s 2ms/step - loss: 0.6841 - accuracy: 0.5527
 Epoch 83/100
 119/119 [-----] - 0s 2ms/step - loss: 0.6841 - accuracy: 0.5527
 Epoch 84/100
 119/119 [============= ] - 0s 3ms/step - loss: 0.6845 - accuracy: 0.5527
 Epoch 85/100
 119/119 [================== ] - 0s 2ms/step - loss: 0.6865 - accuracy: 0.5527
 Epoch 86/100
 Epoch 87/100
 119/119 [=========================== - 0s 2ms/step - loss: 0.6846 - accuracy: 0.5527
 Epoch 88/100
 119/119 [=========== ] - 0s 2ms/step - loss: 0.6832 - accuracy: 0.5527
 Epoch 89/100
 119/119 [=========== - 0s 2ms/step - loss: 0.6867 - accuracy: 0.5527
 Epoch 90/100
 119/119 [============ ] - Os 3ms/step - loss: 0.6826 - accuracy: 0.5527____
 Enoch 91/100
) 119/119 [------] - 0s 2ms/step - loss: 0.6832 - accuracy: 0.5527 ↑ ↓ ⇔ 🗏 💠 🖟 🗎 🗎
 Epoch 92/100
 119/119 [========== ] - 0s 2ms/step - loss: 0.6835 - accuracy: 0.5527
 Epoch 93/100
 Epoch 94/100
 119/119 [============ ] - 0s 2ms/step - loss: 0.6813 - accuracy: 0.5527
 Epoch 95/100
 119/119 [=========== ] - 0s 2ms/step - loss: 0.6828 - accuracy: 0.5527
 Epoch 96/100
 119/119 [=========== - 0s 2ms/step - loss: 0.6804 - accuracy: 0.5527
 Epoch 97/100
 119/119 [============ ] - 0s 2ms/step - loss: 0.6826 - accuracy: 0.5527
 Epoch 98/100
 Epoch 99/100
 Epoch 100/100
 <keras.callbacks.History at 0x7f655d4cf490>
```

# Milestone 4: Model Deployment

Save the best model:

```
#TASK 5 [MODEL DEPLOYMENT]

**Save the best model

import pickle

pickle.dump(knn,open("placement.pkl",'wb'))

model = pickle.load(open('placement.pkl','rb'))
```

#### Integrate with Web Framework

