Predicting Student Success in Government Exam Using Machine Learning

D.Angeline, L.Basti Monisha, Dr.P.Meenakshi Sundari, Msc., M.Phil., SET,Ph.D.

Department of Computer Science, Kamaraj University, Fatima College, Madurai

ABSTRACT
The number of aspirants preparing for government exams is increasing rapidly. Whether studying at a coaching center or through self-study, candidates invest significant effort to succeed. This study aims to predict a student's likelihood of success based on data collected from both successful candidates and those who retake exams.This study achieves an accuracy of 85% in predicting exam success based on student study habits preparation level, and attempt history.

I.   INTRODUCTION

Government exams are known for their difficulty and competitiveness, attracting a large number of aspirants each year. Despite these challenges, many students dedicate themselves to rigorous preparation in pursuit of success. Over the years, the participation rate in these exams has significantly increased. In 1990, the attendance rate for government exams was approximately 20-30%. However, in the 2020s, this figure has risen to 70-90%, marking a three- to fourfold increase.

As competition continues to grow, effective study strategies become crucial. Many students struggle to determine whether self-study or coaching will enhance their chances of success. According to data collected from aspirants, a majority rely on self-study; however, many lose motivation after multiple failed attempts.

This study aims to predict a student's likelihood of success in government exams by analyzing key factors such as study hours, preparation level, number of attempts, and educational background. Machine learning techniques, particularly logistic regression, are employed to evaluate historical data and provide predictive insights. Logistic regression is chosen for its robustness and interpretability in binary classification problems, making it a suitable approach for this study.

DATA COLLECTION

The data for this study were collected using **Google Forms**. aspirants were requested to complete an online survey to help predict the result. The questions included are:
  i. What is your highest level of education?
 ii. What is your current level of preparation for the government exam?
iii. Which government exam are you preparing for?
 iv. How many hours do you study each day?
  v. How many attempts have you made in the government exam?
 vi. Are you preparing for the exam through self study or coaching center?

DATA PREPROCESSING

This study utilizes historical data from government exam aspirants to predict their likelihood of success. The preprocessing phase ensures that the dataset is clean, structured, and ready for training the predictive model.

4.1 Handling Missing Data
```
   categorical_columns = standardized_columns[:-1]  # Exclude target column
 for col in categorical_columns:
   df[col].fillna(df[col].mode()[0], inplace=True)
```

4.2 Encoding Categorical Data
```
    label_encoders = { }
   for col in categorical_columns:
```

```
le = LabelEncoder()
df[col] = le.fit_transform(df[col].astype(str))
label_encoders[col] = le
```

4.3 Target Variable Transformation
```
df["will pass"] = df["will pass"].map({'yes': 1, 'no': 0})
```

4.4 Splitting Data for Training and Testing
```
X = df.drop(columns=["will pass"])
y = df["will pass"]  # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=4)
```

4.4 Feature Scaling
```
from sklearn.preprocessing import MinMaxScaler
```

ANALYSIS
A logistic regression model is trained to classify students as pass or fail based on their study habits
```
model = LogisticRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
```

The model achieves an accuracy of approximately 85%, suggesting that study patterns significantly influence success. Studying with more study hours and fewer attempts has a higher chance of passing.

OUTPUT

Candidate Detail
education: Bachelor degree
preparation: Intermediate preparation (3-6 months)
government exam: SSC CGL
attempt: Second attempt
study hours each day: 3 to 4 hours
Select Option: Self-Study Evaluation
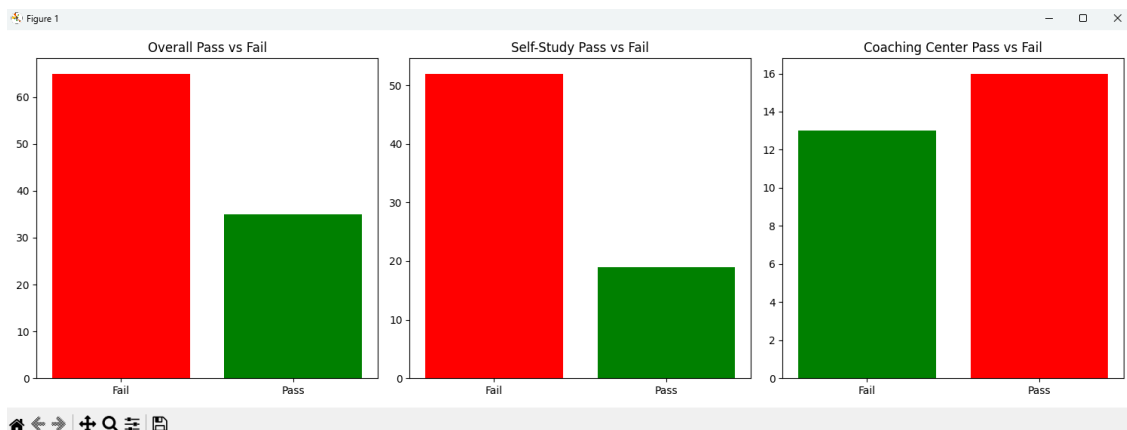
II. CONCLUSION

This study predicts government exam success using machine learning future research will explore deep learning models and additional features, such as past exam scores and psychological factors, to improve prediction accuracy. A web-based application integrating this model could serve as a valuable tool for aspirants.