Loan Approval Risk Prediction Project

Course: NLP (Semester 6) - Pillai College of Engineering

Project Overview:

Loan approval prediction is a crucial aspect of financial decision-making for banks and lending institutions. Automating this process using machine learning techniques helps reduce risks associated with loan defaults while ensuring fair and efficient approval mechanisms. The proposed system leverages historical loan data, applicant financial records, and credit scores to classify loan applications as either approved or denied.

You can learn more about the college by visiting the official website of Pillai College of Engineering.

Acknowledgements:

We would like to express our sincere gratitude to the following individuals:

- **Theory Faculty**:
  - **Dhiraj Amin**
  - **Sharvari Govilkar**
- **Lab Faculty**:
  - **Dhiraj Amin**
  - **Neha Ashok**
  - **Shubhangi Chavan**

Their guidance and support have been invaluable throughout this project.

Project Title: Loan Approval Risk Prediction Project

Project Abstract:

Natural Language Processing (NLP) can play a pivotal role in automating loan application classification by analyzing unstructured financial and personal data. This study explores an NLP- based approach to classify loan applications as likely to be approved or denied by extracting meaningful insights from textual data such as financial history, credit reports, and applicant statements. By leveraging techniques like sentiment analysis, entity recognition, and text embeddings, the model enhances traditional credit assessment methods. Automating this process with NLP not only improves efficiency but also helps lenders make more informed and fair lending decisions while minimizing the risk of defaults.

Algorithms Used:

- **Machine Learning Algorithms**:
    - **Logistic Regression**
    - **Support Vector Machine (SVM)**
    - **Random Forest Classifier**

- **Deep Learning Algorithms**:
    - **Convolutional Neural Networks (CNN)**
    - **Bidirectional Long Short-Term Memory(BILSTM)**
    - **Long Short-Term Memory (LSTM)**
    - **Gated Recurrent Unit(GRU)**
    - **Multilayer Perceptron(MLP)**

- **Language Models**:
    - **ROBERTA(Robustly Optimized BERT Approach)**
    - **BERT (Bidirectional Encoder Representations from Transformers)**

Comparative Analysis:

The comparative analysis of different models highlights their effectiveness in classifying news articles into the correct category. The following table summarizes the **accuracy**, **precision**, **recall**, and **F1-score** of the models tested:

Models	Notes/Predictions
Logistic regression	High accuracy for TF-IDF, struggles with imbalanced data.
Svm	Performs well with TF-IDF, sensitive to feature scaling.
Random forest	High accuracy, handles imbalanced data better.
CNN	Strong with word embeddings but overfits on small datasets.
BILSTM	Slightly better with embeddings, but still weak.
GRU	Performs similarly to LSTM but slightly better in efficiency.
MLP	Works well on structured loan data, but lacks deep contextual understanding
TRANSFORMER	Struggles with small datasets, requires large training data.
BERT	Predictions: (Rejected, Approved, Rejected) for three input texts.
ROBERTA	Predictions: (Rejected, Rejected, Approved) for three input texts.

Conclusion:

In an NLP-based loan approval project, the choice of model depends on the complexity of text data, accuracy requirements, and computational resources. Traditional machine learning models like Logistic Regression and Random Forest can be effective for structured text analysis, such as analyzing financial documents and customer profiles. However, for deeper insights, deep learning models like LSTMs and CNNs can capture complex patterns in application forms, credit history, and customer interactions. If the project involves advanced text understanding, such as extracting sentiments from applicant statements or analyzing unstructured financial documents, transformer-based models like BERT or GPT can provide superior accuracy. While these models improve decision-making, they require significant computational power.

This version tailors the description to focus on **Loan Approval project** while maintaining the structure and content related to the NLP project. It includes relevant details on algorithms used, a comparative analysis table, and the necessary acknowledgments for faculty members involved in the course.