Classification Project on Machine Learning

Implement various ensemble techniques to predict license status for a given business.

Submission Guidelines: Best Practice & Procedures

* The python file should be submitted in IPYNB format
* A PPT presentation outlining your findings must be included in the submission
* To submit the files, create a folder name with your email ID and use underscore for the project subject you have chosen i.e. shashank@gmail.com\_python
* For example: you can refer the project submission demo
* Finally upload the folder containing your files to the designated Batch folder on G-Drive
* For submission kindly watch the attached recording

Description

**Overview**

In machine learning, Classification is one of the most widely used techniques with various applications. For sentiment analysis, spam detection, risk assessment, churn prediction, and medical diagnosis classification have served as a very simple yet powerful method. In this project, we aim to give you hands-on experience and theoretical explanations of various ensemble techniques. You can find the first project of this series.

**Aim**

Understanding various Ensemble techniques and implementing them to predict license status for the given business.

**Data Description**

The dataset used is a licensed dataset. It contains information about 86K different businesses over various features. The target variable is the status of the license, which has five different categories.

**Tech Stack**

* Language: Python
* Libraries: pandas, scikit\_learn, category\_encoders, NumPy, seaborn, matplotlib, xgboost

**Project Approach**

1. Data Description
2. Exploratory Data Analysis
3. Data Cleaning
   1. Missing Value imputation
   2. Outlier Detection
4. Data Imbalance
5. Data Encoding
6. Model Building
   1. Random Forest
   2. AdaBoost
   3. XGBoost
7. Feature importance
8. Hyperparameter tuning
   1. Random search optimization
   2. Grid search optimization
   3. Bayesian optimization

**Note: Perform all the steps to complete the project**

Here are some points you could consider including in your PowerPoint presentation for your license data project, based on the machine learning models you've built:

1. Introduction: Start with an introduction to your project, including the problem statement and the motivation for using machine learning to address it. Explain the dataset and any preprocessing you did before building the models.
2. Exploratory Data Analysis (EDA): Include visualizations and descriptive statistics that summarize the main characteristics of the data. Highlight any interesting insights or patterns you discovered during EDA.
3. Feature Engineering: Describe the features you selected and engineered for the machine learning models. Explain how you decided which features to use and how you transformed or encoded them to work with the models.
4. Model Building: Explain the three models you built - XGBoost, AdaBoost, and Random Forest - and how you selected hyperparameters for each one. Include evaluation metrics and cross-validation results for each model.
5. Model Comparison: Compare the performance of the three models and explain which one performed best for your specific problem. You could include visualizations that show the differences in performance between the models, such as ROC curves or precision-recall curves.
6. Interpretation of Results: Discuss the insights gained from the machine learning models and how they could be used to inform decisions or take action. Highlight any particularly important features that were identified as important by the models.
7. Limitations and Future Work: Discuss any limitations of the models or the dataset, and suggest areas for future work or improvement. This could include collecting additional data, using more sophisticated machine learning techniques, or testing the models on different datasets or problem domains.
8. Conclusion: Summarize the key findings of the project and the impact they could have. Emphasize the importance of the problem and the value of using machine learning to address it.

Remember to include relevant visuals, such as tables, charts, and graphs, to help illustrate your points and make the presentation more engaging. Also, keep in mind your target audience and adjust the level of technical detail and jargon accordingly.