Machine Learning Model for Real-life Predictions

Student Name

2/18/23

Table of contents

List of Figures

List of Tables

Preface

Write your content.

Abstract

Type here the abstract of the project.

1 Introduction to Project

See this how you can add a reference, Knuth (1984).

1.1 Overview

1.2 ExistingSystem

1.3 Objectives of Project (Must be clearly, precisely defined and must be covered in the work)

2 Pre-Processing and Exploratory Data Analysis

Once you have your dataset, you need to preprocess it to make it suitable for machine learning. This includes cleaning the data, handling missing values, encoding categorical variables, and scaling numerical features. This step is crucial as the accuracy of your model will depend on the quality of your data.

2.1 Dataset Collection

2.2 Data Pre-processing

Some issues to consider to be included: - Are there missing values in your data? Is there any pattern in whether or not a particular value is missing? How will the missing values be dealt with?

- Are there problems because of data size (rows or columns)? Are some algorithms more affected? Can you come up with any solutions?
- If you are doing any clustering, are there clusters in the data? How many? How well separated are they? Do different methods agree on the number of clusters?
- Is scaling of the variables an issue? For what methods?
- Are there outliers in the data? How did you deal with them?
- Does transforming the data simplify any analysis?

2.3 Exploratory Data Analysis and Visualisations

Some of the steps, included here:

- Data visualization: Data visualization is the process of creating graphs and charts to help you understand the patterns and trends in the data. This step can help you identify outliers and anomalies in the data.
- Statistical analysis: Statistical analysis involves applying statistical methods to the data to identify patterns, trends, and relationships. This step can help you identify correlations between variables and understand the distribution of the data.

- Feature selection: Feature selection is the process of selecting the most important variables that will be used in the machine learning model. This step can help you identify which variables are most predictive and which variables can be ignored.
- Dimensionality reduction: Dimensionality reduction is the process of reducing the number of variables in the data set. This step can help you reduce the complexity of the model and improve its performance.

2.4 Other Related Sections (Optional)

3 Methodology

3.1 Introduction to Python for Machine Learning

3.2 Platform and Machine Configurations Used

Such Google Colab, Kaggle, Syzygy, or your own machine then the machine configuration.

3.3 Data Split

Data splitting is the process of dividing the data into training, validation, and test sets. This step can help you evaluate the performance of the model on new data.

3.4 Model Planning

The next step is to select the appropriate machine learning or deep learning models (2 - 3) that can be used to solve the problem. This involves evaluating different models and selecting the one that is best suited for the problem.

3.5 Model Training:

This involves using the training data to adjust the model parameters to improve its accuracy.

3.6 Model Evaluation

Once the model is trained, the next step is to evaluate its performance on a validation dataset.

3.7 Model Optimization

Based on the results of the model evaluation, the next step is to optimize the model to improve its performance. This involves making changes to the model's architecture, tuning the hyperparameters, and retraining the model.

3.8 Final Model Building

Once the model is optimized, the final step is to fit the model and report final test error for each model

Present the information with support Screen Shots/ Figures (Each Figure must be numbered and Description of Figure must be provided)

4 Results

To compare different machine learning models in the results section, here are some steps to consider:

4.1 Description of the Models

Start by providing a brief description of each model, including its purpose, algorithm type, and any relevant parameters or features used.

4.2 Performance Metrics

Describe the performance metrics used to evaluate the models, such as RMSE, MSE, accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC), etc based on the model class.

4.3 Results Table

Create a results table that presents the performance metrics for each model in a clear and concise format. The table should include the name of the model, the performance metric, and the value of the metric. It can also be helpful to include the standard deviation or confidence intervals for each metric.

4.4 Interpretation of the Results

Provide an interpretation of the results, highlighting any significant differences between the models. Discuss which model performed the best and why, as well as any limitations or potential sources of error.

4.5 Visualization

Provide visualizations to support the interpretation of the results, such as bar charts or box plots. These can help to illustrate the differences between the models and make the results more accessible to readers.

4.6 Sensitivity Analysis

Conduct a sensitivity analysis to test the robustness of the models. This can involve varying the parameters or features used in the models and re-evaluating their performance.

5 Conclusion

Conclude by summarizing the key findings of the analysis, discussing the implications of the results, and suggesting areas for future research.

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.