Tamader AboAlhassan - Afrah Almoutiri - Moneera Alfulaij
2/11/2021

أكاديمية سدايا
SDAIA Academy

Classification Analysis of Mushrooms for California Garden

Abstract

California Garden is a pioneer in packaged ambient foods across the world. It offers a large product portfolio of high-quality convenient foods, California Garden committed to a philosophy of producing premium quality canned foods based on the meticulous selection of the world's highest quality raw materials. California Garden decided to produce canned mushrooms, it needed a classification of mushrooms to help it choose the best types of mushrooms. Therefore, a study was conducted to bring out the best types of mushrooms that are suitable for human consumption.

Design

This project originates from the Data Science Bootcamp (T5) to Comparison of classifier Algorithms of mushrooms, based on many features. through the Logistic Regression, K-Nearest Neighbor, Naive Bayes, Decision Tree, SVM, Random Forest, GaussianNB and XGBClassifier .

Data

Using the Classification Analysis of Mushrooms from kaggle.com. The dataset contains 8125 rows and 23 features: Class, cap-shape, cap-surface, cap-color, bruises, odor, gill-attachment, gill-spacing, gill-size, gill-color, stalk-shape, stalk-root, stalk-surface-above-ring, stalk-surface-below-ring, stalk-color-above-ring, stalk-color-below-ring, veil-type, veil-color, ring-number, ring-type, spore-print-color, population and habitat.

Algorithms

The methodology used in this project is: Problem understanding, Data validation, Data exploration, Data visualization, Feature engineering, Feature selection, Training and modeling the data.

Tools

- Numpy and Pandas for data manipulation.
- Matplotlib and Seaborn for plotting.
- LogisticRegression model from sklearn.linear_model class to build a classification algorithm that is used to predict if the client will subscribe.
- train_test_split function in Sklearn model selection for splitting data.
- KNeighborsClassifier model from sklearn.neighbors to build a classification algorithm that is used to predict if the client will subscribe.
- DecisionTreeClassifier model from sklearn.tree to build a classification algorithm that is used to predict if the client will subscribe.
- RandomForestClassifier model from sklearn.ensemble to build a classification algorithm that is used to predict if the client will subscribe.
- Measure performance of each algorithm using precision_score, recall_score, accuracy_score, roc_auc_score and confusion_matrix from sklearn.metrics module.
- Show the report about the data using ProfileReport from pandas_profiling.
- Jupyter notebook to execute the code.
-

Communication

Presentation.