|  |  |
| --- | --- |
| Internship Project Title | RIO-125: Rank features of Smartphone |
| Project Title | Rank Features of a Smartphone - Build a Python Application to Classify and Rank Dataset |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | DEBASHIS ROY |
| Name of the Institute | MEPCO SCHLENK ENGINEERING COLLEGE |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 12/01/2022 | 24/03/2022 | 60 | Google Collab, Jupyter network | Python, MS Word, Sklearn, app sheet |
| **Project Synopsis:**  This project focuses on implementing a classifier using machine learning algorithms to extract features of a smartphone. This project approach allows to give a grade to the various features based upon price range and battery power. From the evaluation of this project, it can be concluded that the proposed machine learning techniques and the analysis are effective and practical methods for predictive and descriptive analysis. | | | | |
| **Solution Approach:**  By using the scikit-learn library we train the model to identify the mobile by using their id of textual data on the ranking features of smartphone. | | | | |
| **Assumptions:**  Some columns in the data set are dropped which is not useful for the sentiment analysis and the values in the columns are assigned as true. | | | | |
| **Block Diagram:**    **DATA ANALYSIS AND RANKING**  **DATA VISUALIZATION**  **DATA CLASSIFICATION**  **DATA PREPROCESSING** | | | | |
| **Algorithms:**  Some classification algorithm like KNN, SVM, Perceptron are used. | | | | |
| **Data Preprocessing:**  It is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms. Here the features of smartphone has taken and the data is processed by handling the missing values and described the data types of each attribute. | | | | |
| **Data Classification:**  KNN: k-NN is a type of classification where the function is only approximated locally and all computation is deferred until function evaluation.  Perceptron: The perceptron is an algorithm for supervised learning of binary classifiers. A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class. It is a type of linear classifier.  SVM: Support Vector Machine is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification of this data.  Logistic Regression: Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model.  GaussianNB: A Gaussian Naive Bayes algorithm is a special type of NB algorithm. It's specifically used when the features have continuous values like battery power. It's also assumed that all the features are following a gaussian distribution.  Random Forest: Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.  Decision Tree: A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. | | | | |
| **Outcome:**  The output of the project is done by naïve bayes, svm, linear regression, perceptron, decision tree, random forest and knn algorithms for the ranking and analysis.  **RANKING THE DATASET**    **ANALYSIS USING DIFFERENT MODELS**    **ACCURACY FOR DIFFERENT MODELS** | | | | |
| **Exceptions considered:**  While preprocessing the data hashtags, punctuations are removed.  Some columns in the data set are also considered which is not useful for the analysis. | | | | |
| **Enhancement Scope:**  In future we can serve the best model for the customers by analyzing the price and battery power of a smartphone. | | | | |
| **Challenges:**   * Completed the Industry Project according to the project guidelines mentioned. * Worked on the Project Report. * Checked out the Digital Discussion Room for more resources and direction to follow on my current situation. * Completed writing the interim report. * Tried increasing the accuracy value. | | | | |
| **Output and screenshots:** | | | | |
| **Link to Code and executable file:**  https://github.com/RajeshwariKarthikeyan/Rank-Features-of-Smartphone | | | | |