

# GOOGLE PLAY STORE DATA ANALYSIS

SAGAR PANDEY 18BCE1002



## INTRODUCTION

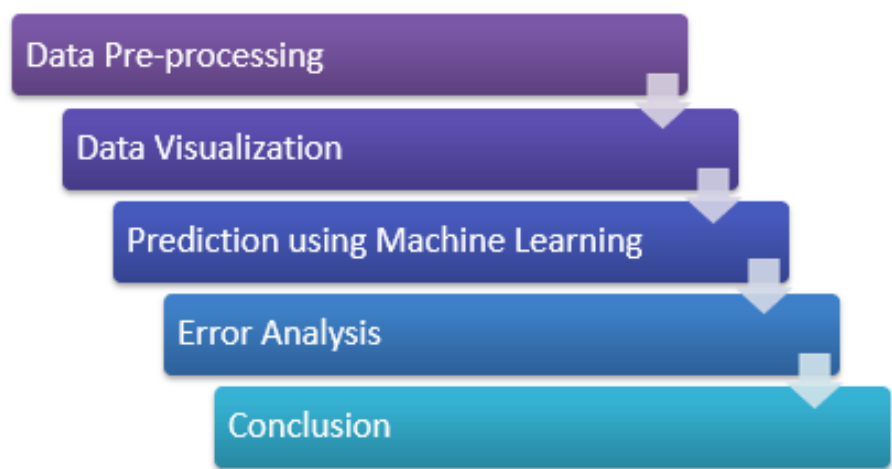
The Internet is a true gold mine of data. E-commerce and review sites are brimming with a lot of untapped data with a prominent potential to convert into meaningful insights that can help with robust decision making. Here, we explore using data science and machine learning techniques on data retrieved from one such avenue on the internet, the **Google Play Store**. The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. The dataset is chosen from Kaggle. It is the web scraped data of 10k Play Store apps for analyzing the Android market. It consists of in total of 10841 rows and 13 columns.

## OBJECTIVES

We will be answering the following questions:

- The main focus of the project is to felicitate the developer to predict the installs based on other factors of the application.
- How are the features related to each other?
- After visualization, which Machine Learning Models will give the best prediction accuracy?
- To generate the error score of the various algorithms like Random tree Regressor, decision regressor tree and KNeighbor Classification.

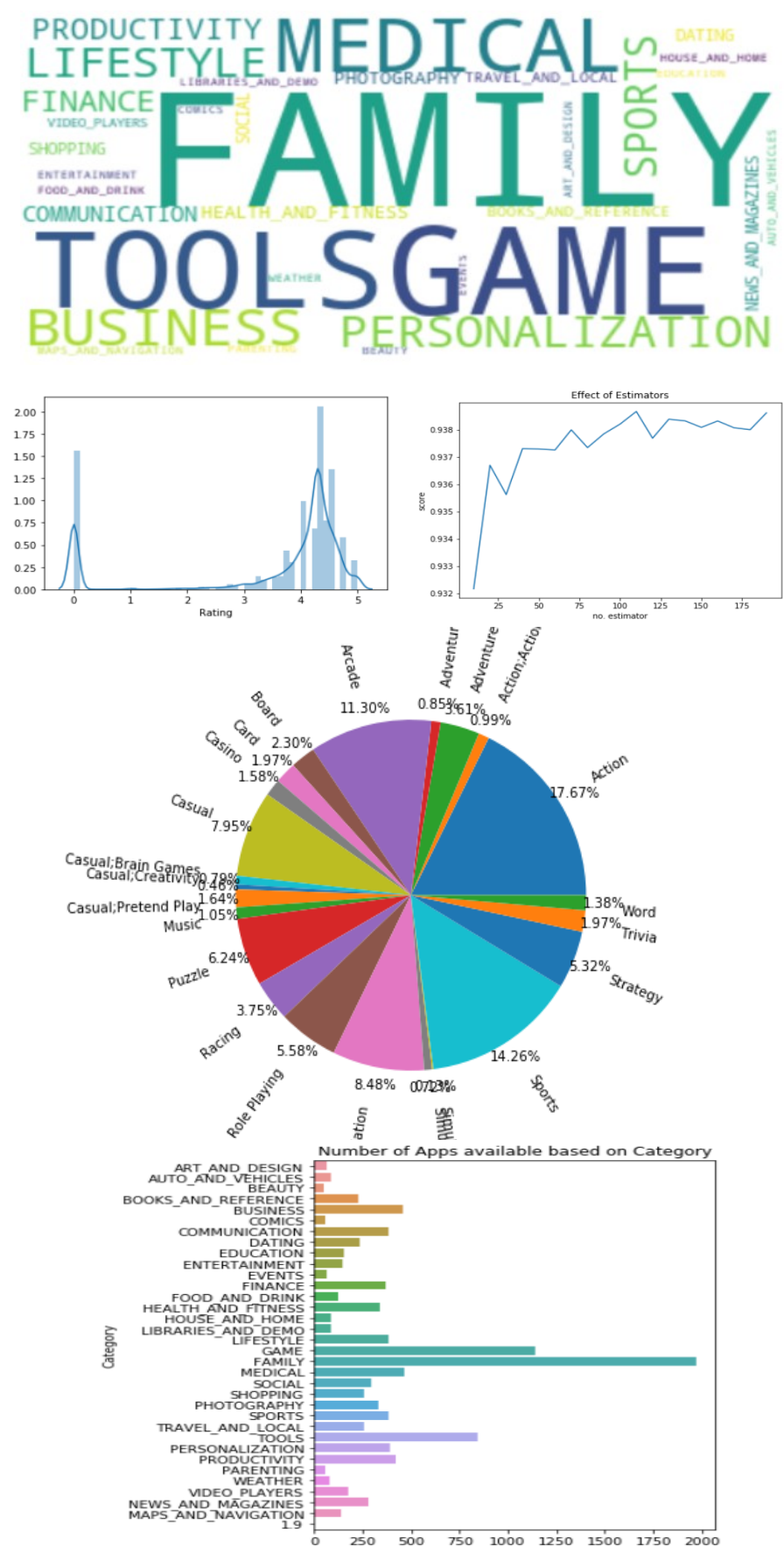
## WORKFLOW



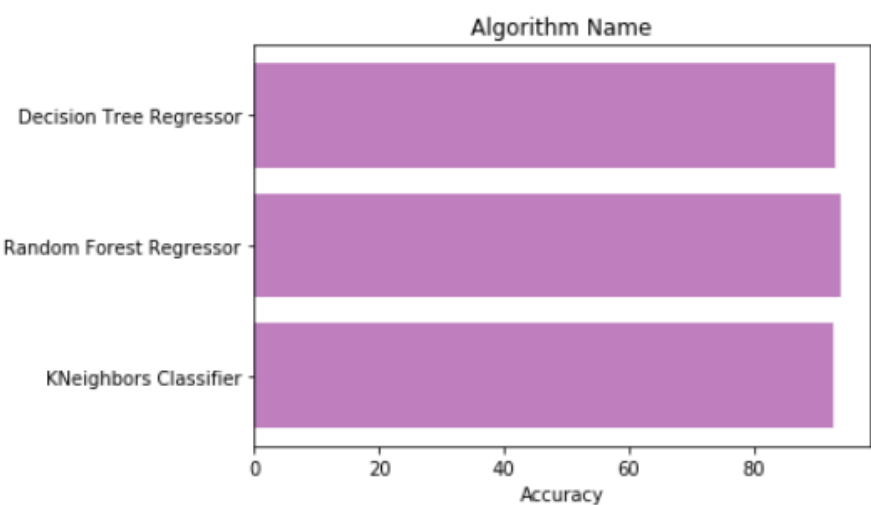
## DESCRIPTION

- 1) App (Name)
- 2) Category (App)
- 3) Rating (App)
- 4) Reviews (User)
- 5) Size (App)
- 6) Installs (App)
- 7) Type (Free/Paid)
- 8) Price (App)
- 9) Content Rating (Everyone/Teenager/Adult)
- 10) Genres (Detailed Category)
- 11) Last Updated (App)
- 12) Current Version (App)
- 13) Android Version (Support)

## DATA VISUALIZATION



## RESULTS



The algorithms namely KNeighbor, Random Forest Classifier and Decision Tree Regressor were used to fit the dataset and predict the results. The **Random Forest Regressor** proved to be the best suited algorithm for this dataset with an accuracy of **93.85%** followed by **Decision Tree Regressor** and **KNeighbor** with accuracies of **93.00%** and **92.59%** respectively.

## CONCLUSION

- The Random Forest regressor gives the best accuracy but lags behind in training time since around 150 trees were made in this process to achieve this accuracy.
- On the other hand, the Decision Tree Classifier proves to be the fastest with quite acceptable accuracy of 93.0%