



## **Megastore Profit Prediction**

### **2: classification**

Team ID	SC_31
---------	-------

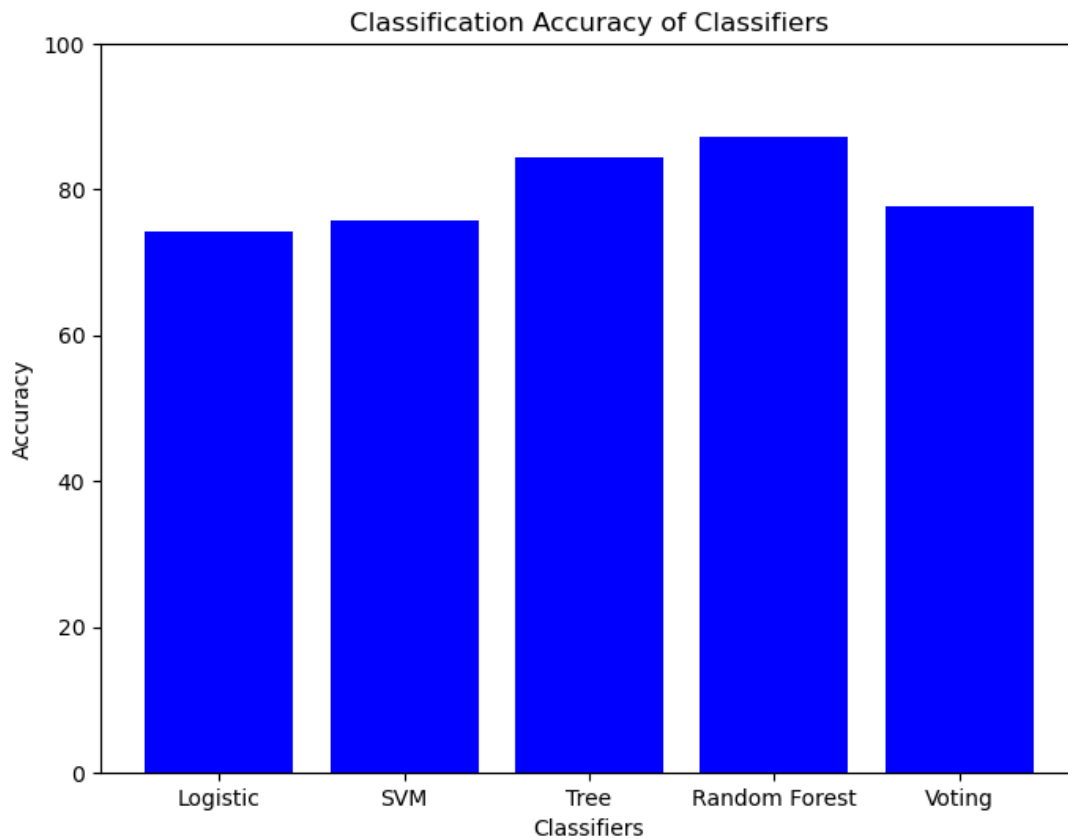
Name	ID
محمد محمود فوزي خليل	20201700736
فتحي ناصر فتحي	20201700575
مازن عبداللطيف لطفي عبد اللطيف	20201700642
احمد ياسر محسن صالح	20201700095
محمد ابراهيم عبدالحميد محمد	20201700651
محمد عبد الله عبد الحكيم سالم	20191700835

## The Report

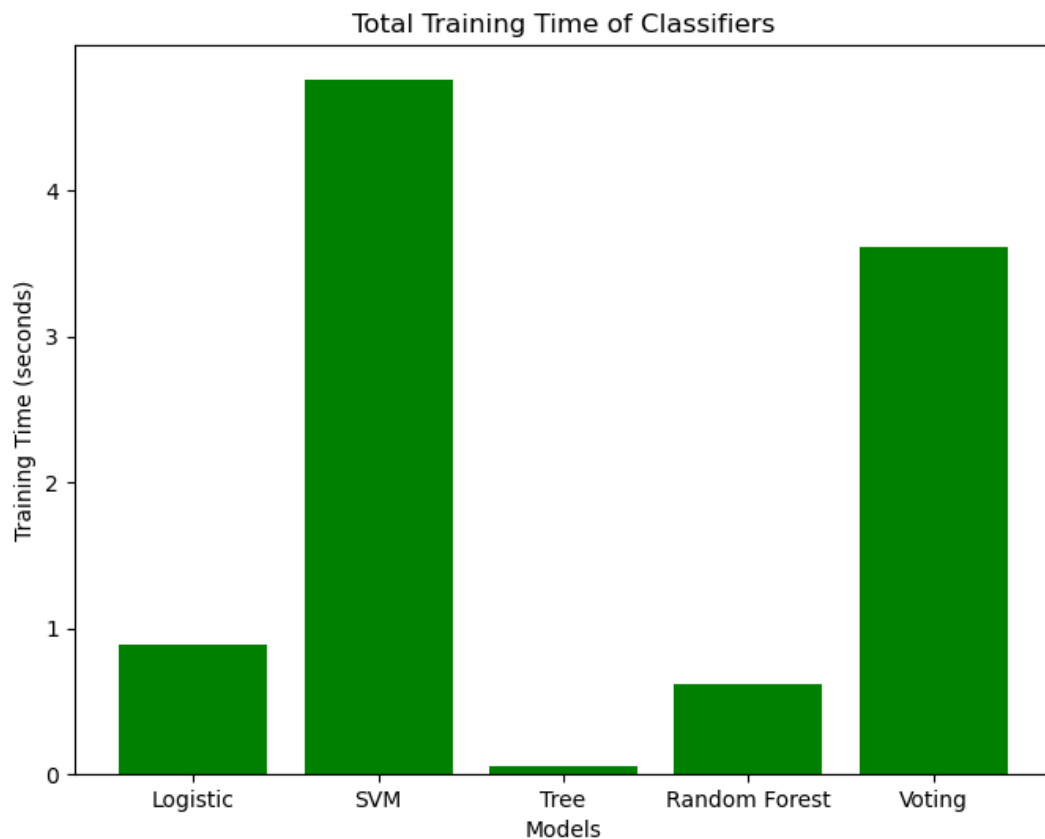
---

1.The classification accuracy is under range from 60% to 88 % whenever we change the type of classification or we change the hyperparameter of any classification and this is evident in his bar graph.

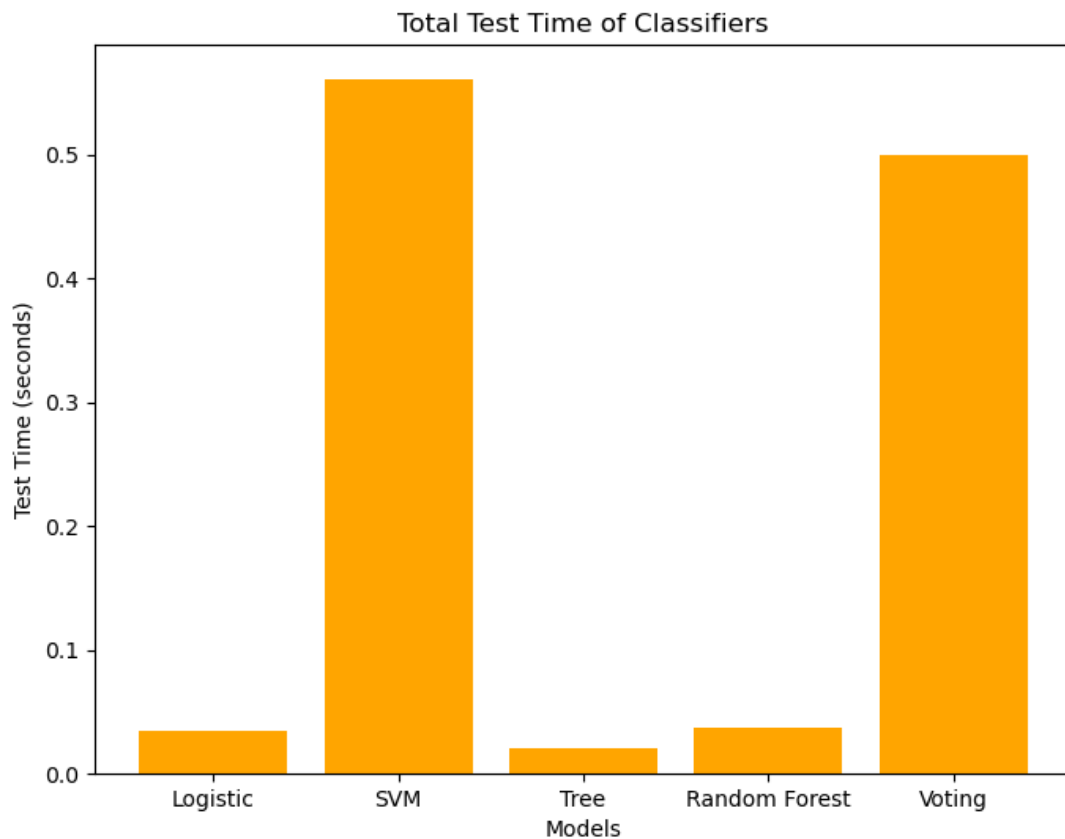
---



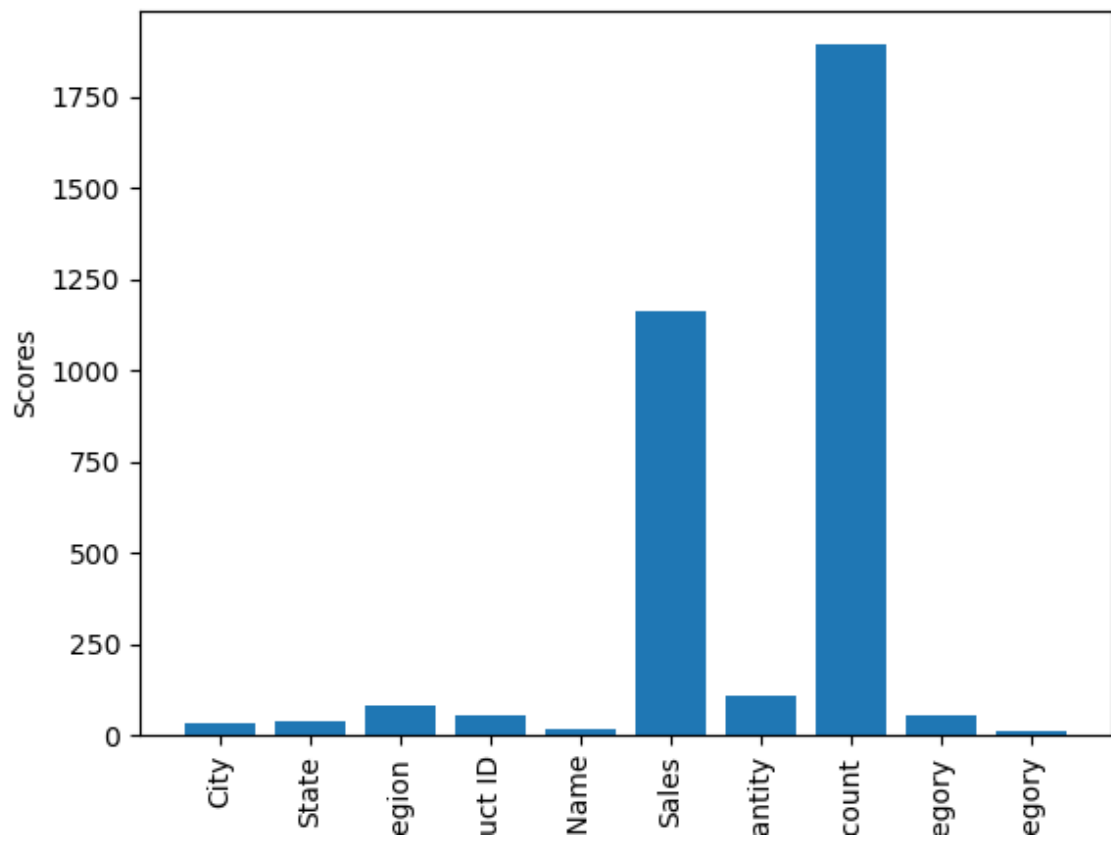
2. We used the time before and after each fit to calculate the total training time. We notice that it changes significantly with each classification and it changes from classifier to classifier and also changes with the change of the hyperparameters of each classifier.



3. We used the time before and after each predict to calculate The total testing time , We notice that it changes significantly with each classification and It changes from classifier to classifier and also changes with the change of the hyperparameters of each classifier.



4. In the classification phase, the feature selection process may differ from the previous regression phase. In regression, the feature selection process involved using correlation analysis to identify relevant features. However, in classification, a different approach was adapted. For classification, the feature selection process utilized the K-best feature selection method. This technique selects the K best features based on statistical significance and ability to contribute to the classification task. The specific criteria used to determine the best features may vary, but common approaches include chi squared test, mutual information, or ANOVA F-value. By adopting the K-best feature selection approach in the classification phase, the feature set was refined, focusing on the most informative and discriminative features. This process aimed to enhance the model's ability to accurately classify instances and improve overall classification performance.



## 5: hyperparameter tuning:

Hyperparameter tuning plays a crucial role in optimizing the performance of machine learning models. In the case of classification models like logistic regression and support vector machine (SVM), tuning specific hyperparameters such as 'C' (regularization parameter) and 'kernel' for SVM can have a significant impact on the models' performance.

### 1. Logistic regression:

#### A. Penalty

- L1: 74.148
- L2: 74.35
- None: 74.21

#### B. C

- 1.0: 74.079
- 0.01: 65.948
- 10: 74.287

#### A. Penalty:

L2 regularization (Ridge) achieved the highest accuracy of 74.35%.

L1 regularization (Lasso) achieved an accuracy of 74.148%.

Not applying any regularization (None) resulted in an accuracy of 74.21%.

Among the regularization techniques tested, L2 regularization performed slightly better than L1 regularization and not applying any regularization.

B. C:

The C value of 1.0 yielded an accuracy of 74.079%.

A lower C value of 0.01 resulted in a lower accuracy of 65.948%.

A higher C value of 10 achieved an accuracy of 74.287%.

From the C values tested, a value of 1.0 yielded the highest accuracy, followed by 10. However, a lower C value of 0.01 led to a significant decrease in accuracy.

Overall, the logistic regression model achieved moderate accuracy ranging from 65.948% to 74.35%. The best accuracy was achieved when using L2 regularization with a C value of 10.

## 6: Conclusion

- In this phase of the project, the same preprocessing techniques used in the previous phase were applied. The K-best feature selection method was employed and proved to be effective in selecting the most relevant features for the classification task.
- Multiple classification models, namely Logistic Regression, SVM, Random Forest, and decision tree, and voting algorithm were evaluated, and their performances were compared. The intuition behind this approach was to explore different models with varying characteristics and capabilities, expecting that some models would outperform others in terms of accuracy.
- The hyperparameter tuning process was also conducted, aiming to optimize the performance of each classification



model. By adjusting the hyperparameters, such as regularization parameters for SVM, the models' performance could be further improved.

- Through experimentation and evaluation, conclusions can be drawn about the effectiveness of each classification model. The accuracy of each model was monitored and compared, and it was observed how the hyperparameter tuning influenced the models' performance.