

Group 20: Kobe Bryant

Apoorvarani K Rajashekar, Sidra Aziz, Xiongjun Wang, Anna Losovska

Challenge

0 - Shots Missed

1 - Shots Made

Predict whether Kobe Bryant will make a shot or not.

Dataset

- 6 types of different shots played by Kobe in provided data set.
- 30,697 records.
- 7 categorical features, with 2-8 possible values each. There are also 11 Numerical Features and 4 nominal features in specific.

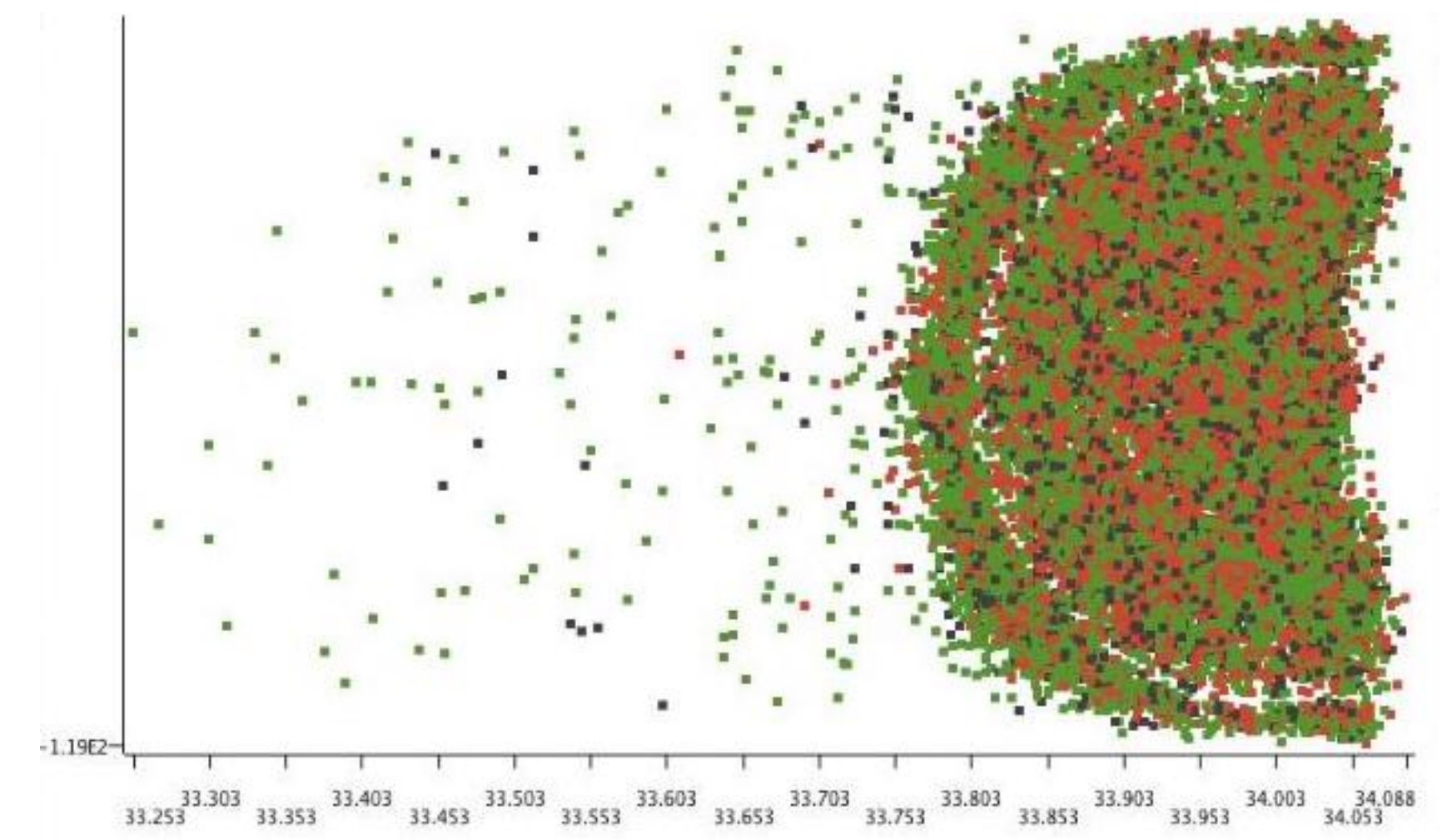


Fig. 1 - Shots made by Kobe from different zones in the court

Data Understanding

- Values were not fully represented for some features.
- Dataset is unbiased with respect to Target Attribute

Data Preparation

- **Selected Attributes:**
 - Shot Made Flag
 - Shot Type
 - Latitude, Longitude
 - Minutes remaining, seconds remaining
 - Shot Zone Area
 - 5,000 missing values

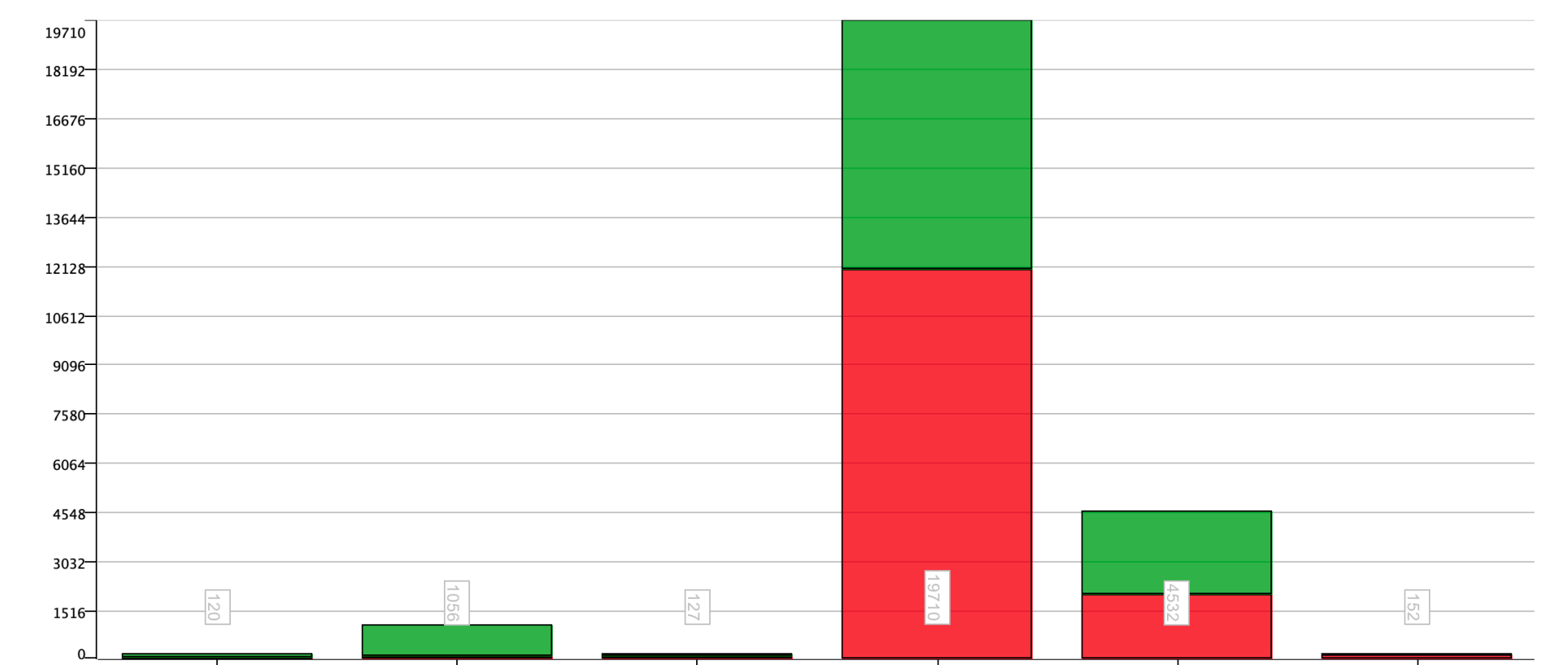


Fig. 2 - Types of Shots made by Kobe, Jumpshot attempts are highest

Confusion Matrix - 4:82:51 - Scorer

File	Hilite
shot_made...	0
0	12547
1	6755

Modeling

- **Final Partitioning:**
 - 80% Training Data
 - 20% Test Data
- **Classification Models used:**
 - Naive-Bayes
 - Random Forest
 - Decision Trees
 - Logistic Regression
- The quality measure is to **minimise Error rate and Increase accuracy** in prediction whether Kobe will make a shot or not.
- **Final Model**
 - Random Forest
 - Information Gain Split criterion

Correct classified: 17,257	Wrong classified: 8,440
Accuracy: 67.156 %	Error: 32.844 %
Cohen's kappa (κ) 0.306	

Fig. 3 - Confusion Matrix for Random Forest

Results and Evaluation

Models	Accuracy
Decision Tree	67.23%
Random Forest	67.15%
Logistic Regression	68.11%
Naive Bayes	61.52%

- Data on left shows the accuracy rate for various models tested on dataset.

Conclusion

Based on reason that Random Forest works good on Categorical Target attribute as compared to Logistic Regression that generally works on continuous class labels, as well as its performance is better as compared to single decision trees, we chose RF as final model.

Knime-Workflow

