

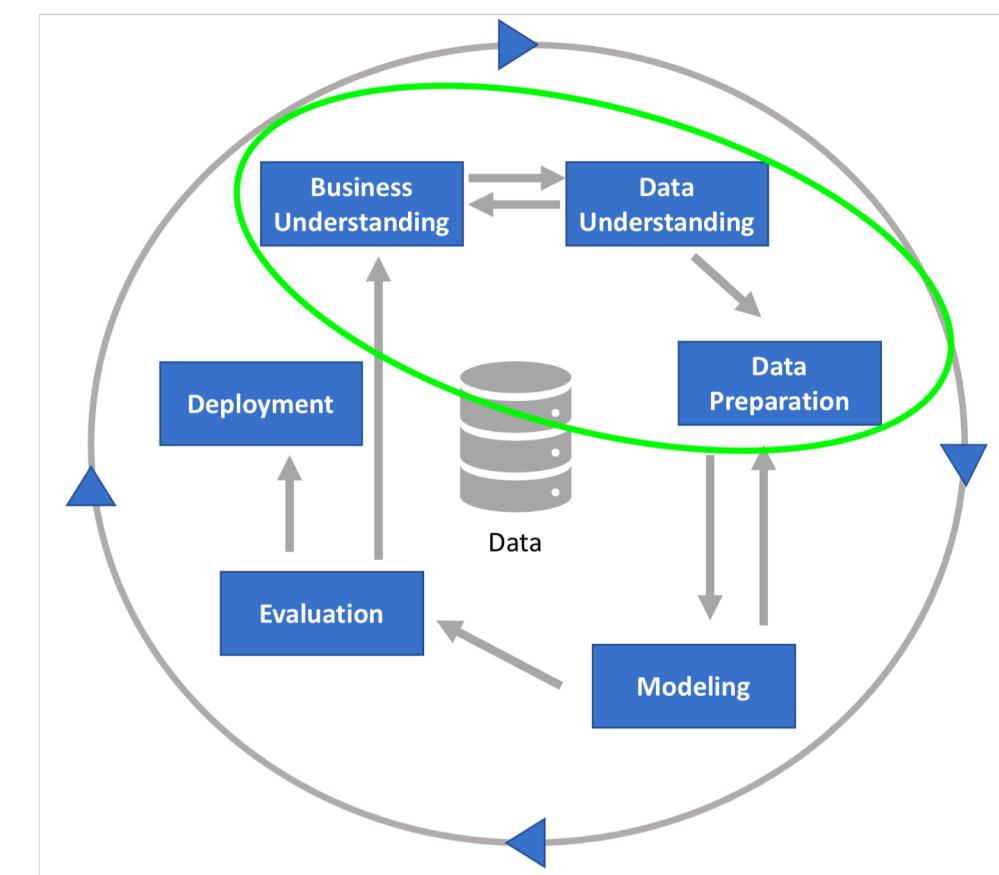
Kickstarter Classification Project



Karan Goda 460496371

Setup

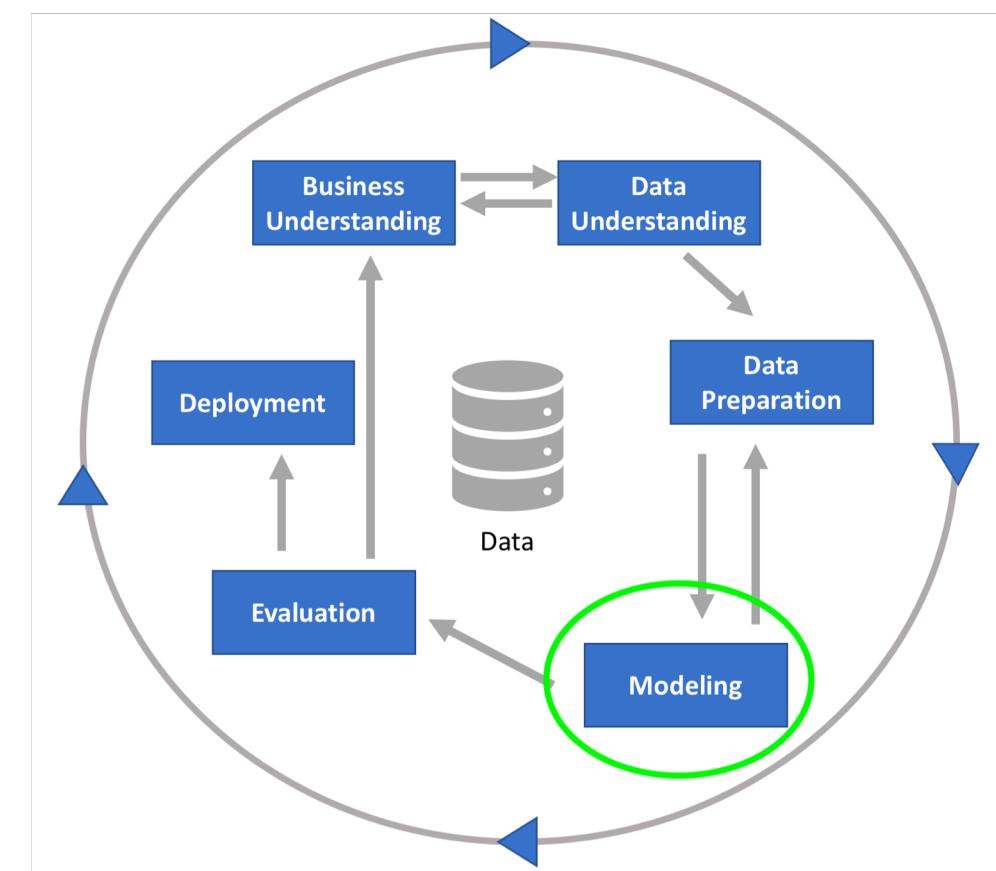
- The dataset is Kickstarter projects and their information.
- Research Question: Are we able to use the project's features such as country or category to predict if the project reaches funding or not.
 - H₀: Baseline values are better than the classifiers.
 - H₁: There is at least one classifier with better results than baseline.
- Effectiveness will be measured using F1 score, precision, accuracy.



We are here in the CRISP-DM cycle

Approach

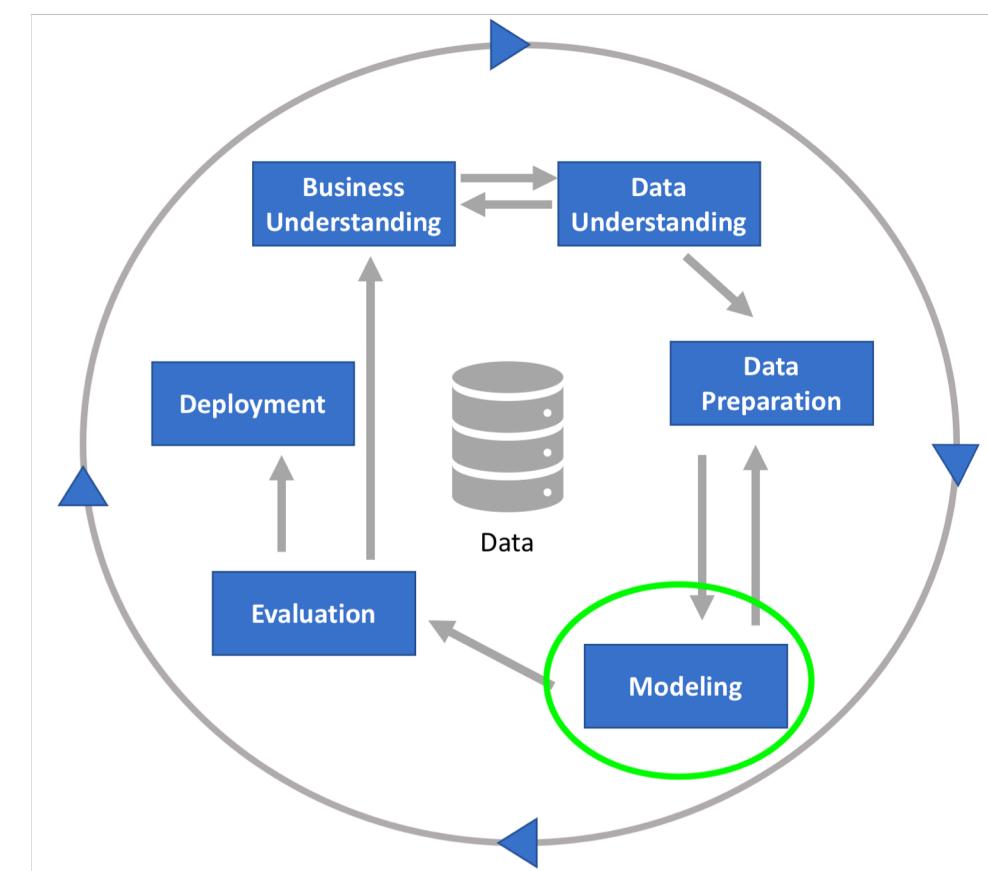
- Supervised methods will be used since datatypes are known
- Models we are using are:
 - Logistic Regression
 - Linear SVC
 - Decision Tree
 - Random Forest
 - KNN



We are here in the CRISP-DM cycle

Experimental Setup

- All the models contains hyperparameters we can tune.
- However for most of the models, we will let Jupyter use the default parameters to create the models.
- The number of estimators parameter in random forest will be modified to 100.
- This increases model effectiveness and decreases overfitting.

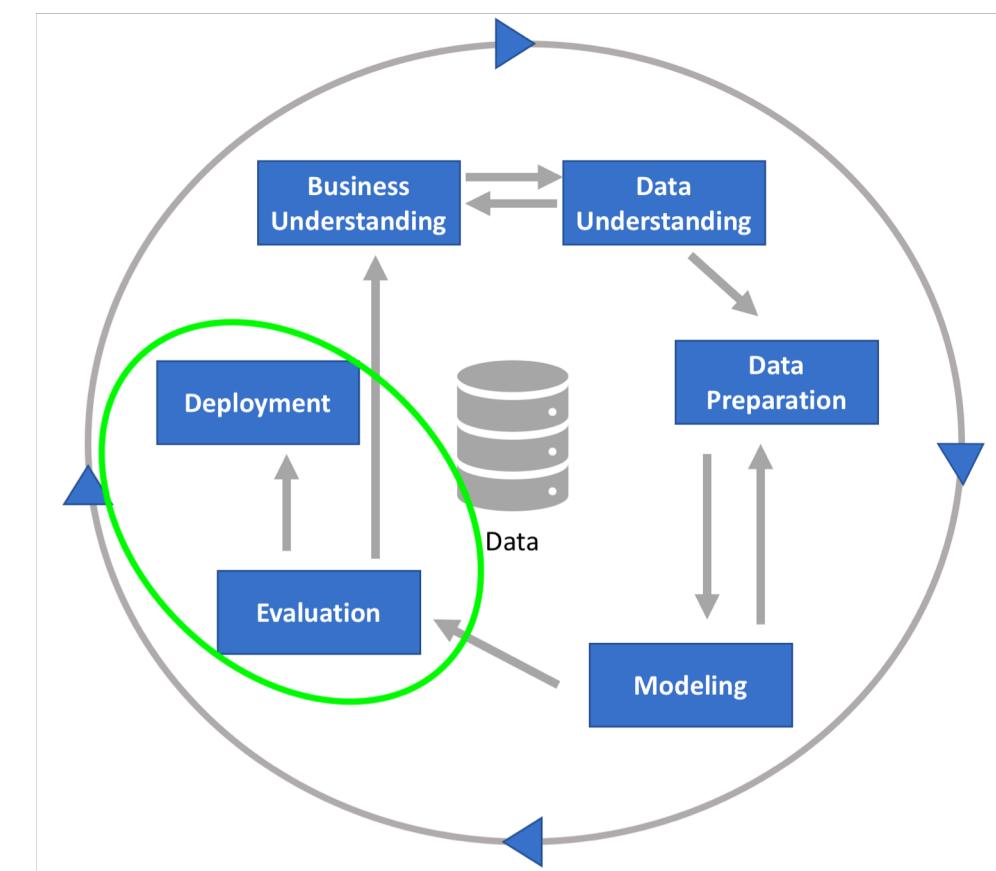


We are here in the CRISP-DM cycle

Results

Model	F1 Score	Accuracy Score	Precision Score
Random Forest	0.579	0.633	0.593
Decision Tree	0.534	0.587	0.540
KNN	0.531	0.600	0.548
Logistic Regression	0.396	0.633	0.816
Linear SVC	0.300	0.361	0.395

- Model Scores are well rounded
- High Performance
- Better than baseline
- Best Model is Random Forest



We are here in the CRISP-DM cycle

Thank You!



KICKSTARTER

