Prophecy - An automated machine learning training and serving platform

Denys Kovalenko, Prabhant Singh (denis.v.kovalenko@gmail.com , prabhant.singh@ut.ee)

Machine Learning is on the rise, and more organizations are starting own data science projects. The major problem in these projects is lack of skilled data science and engineering talent to train and deploy the solution. **Prophecy** helps teams to focus on feature engineering by having automated model training and deployment/serving infrastructure, leveraging best practices from software engineering and Auto ML approaches to allow easy and fast experimentation.

AutoML - Model Training and Selection

The selection of machine learning model and hyperparameter optimization is done by using Genetic programming[1]. After obtaining the optimized algorithms Prophecy also outputs the trained pickle file, python file with the meta-data and visualisations of model performance like ROC Curves and Confusion matrix. This serves as the complete backend system for the user where he can train, evaluate and get the final model as python file for replication.

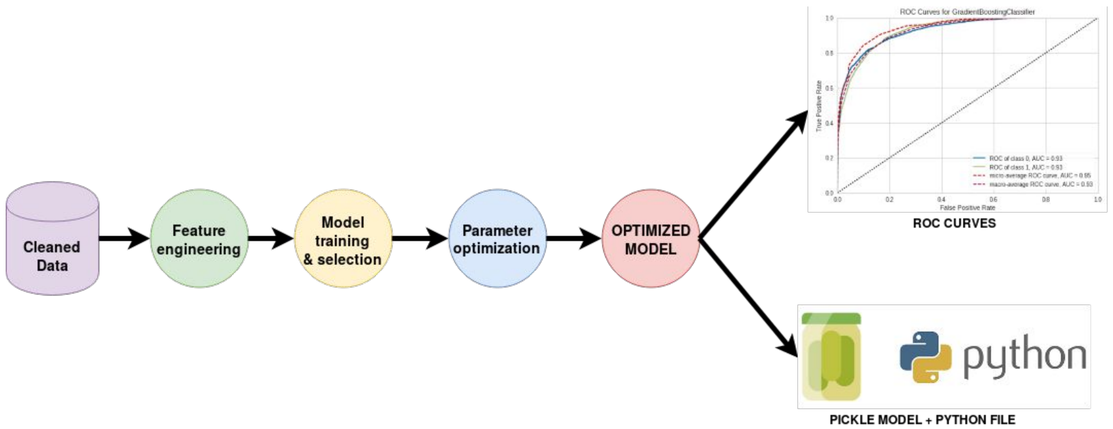


Figure 1: Prophecy's Machine learning Pipeline

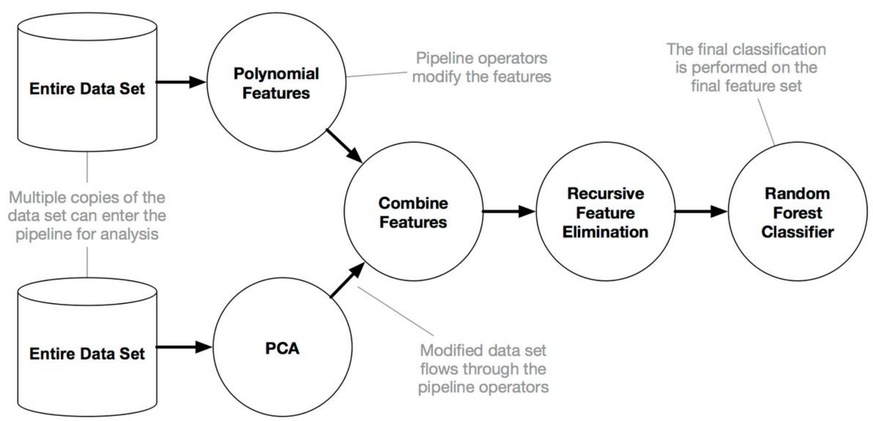
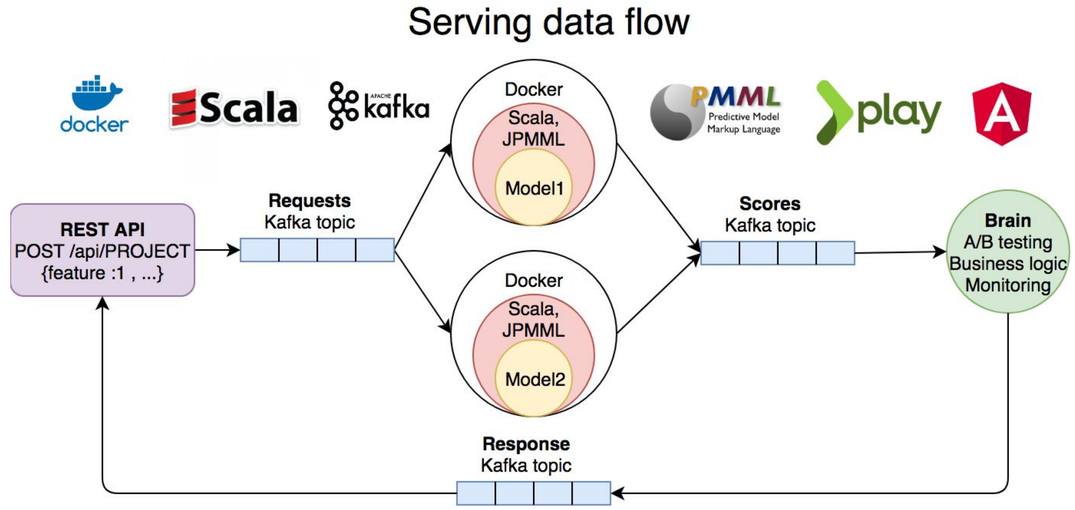


Figure 2: Feature construction Pipeline

Model Serving and Infrastructure

Prophecy deploy best model trained with AutoML and also allows 3-click deployment of separately trained machine learning models into shadow mode on production environment. It enables no-downtime update of champion model, horizontal scalability and A/B testing. New models are converted into PMML and deployed as docker container serving inferences with JPMML, allowing low-latency. Using Kafka as a messaging system allows to persist requests and extend system with Dataset Shift detection etc.



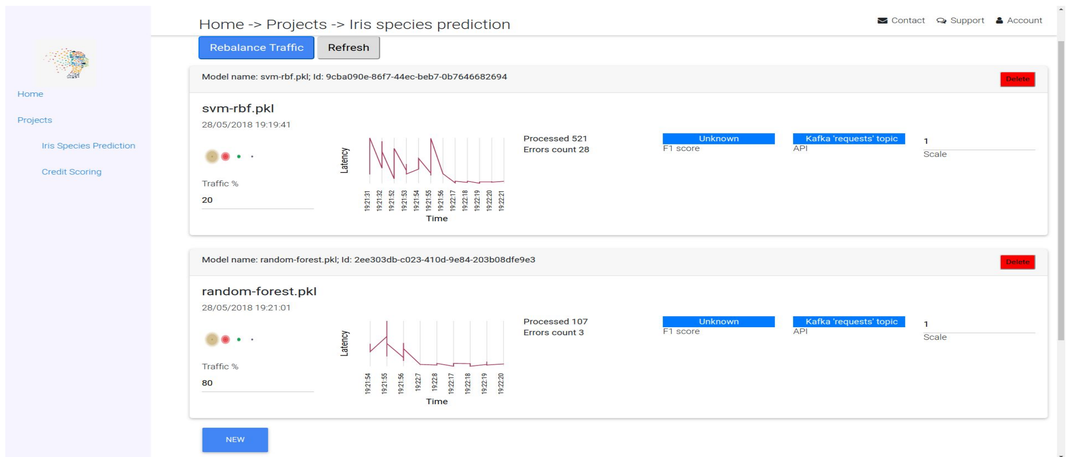


Figure 4: Data Science model dashboard

UNIVERSITY OF TARTU
Institute of Computer Science

Study IT in .ee
sponsored by Skype™

Setup instructions and demo:
https://prabhant.github.io/AutoML/

References:
[1]Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science, R.Olson.
[2] Machine Learning Logistics, Ted Dunning & Ellen Friedman. O'Reilly
[3] Hidden Technical Debt in Machine Learning Systems. Sculley et al.