Enhancing Power of ML Prediction using Knowledge Graph

Zaigham Abbas,Andrea Polini

{zaigham.abbas,andrea.polini}@unicam.it

Computer Science Division, Science and Technology School, University of Camerino

Motivations

Feature Engineering is backbone and core component of Machine learning.

Feature Engineering supports supervised learning,unsupervised learning and Reinforcement learning in the

- Classification
- Regression
- Clustering

Classification is the process of classifying a set of data into distinct categories.classes are often represented as target, label, or categories.Regression is a machine learning strategy that uses an algorithm to predict continuous outcomes.Clustering is the task of dividing a population of data points into groups that are similar to each other and distinct to each other. It is a grouping of items based on their similarity and dissimilarity.

feature selection is hectic and **high complexity** task.the prediction accuracy of ML model may increase in case of optimal feature selection or may decrease otherwise.

Conclusions

We propose a framework to improve the power of ML prediction with automatic feature engineering using knowledge. Using feature engineering techniques, We develope a prototype Automatic Knowledge Graph (AutoKG). This tool builds Knowledge Graph KG in a quick way using the semantics of any Wikipedia Page with Domain Expert Knowledge. we crawl thousands of the wiki web pages,apply web scraping and indexing and formulate the Knowledge Graph relevant to the features or concepts available in the original dataset. A distance matrix technique is applied to the features and selects those similar features that are the minimum distance to the source feature. A new dataset is generated with additional new features.

References

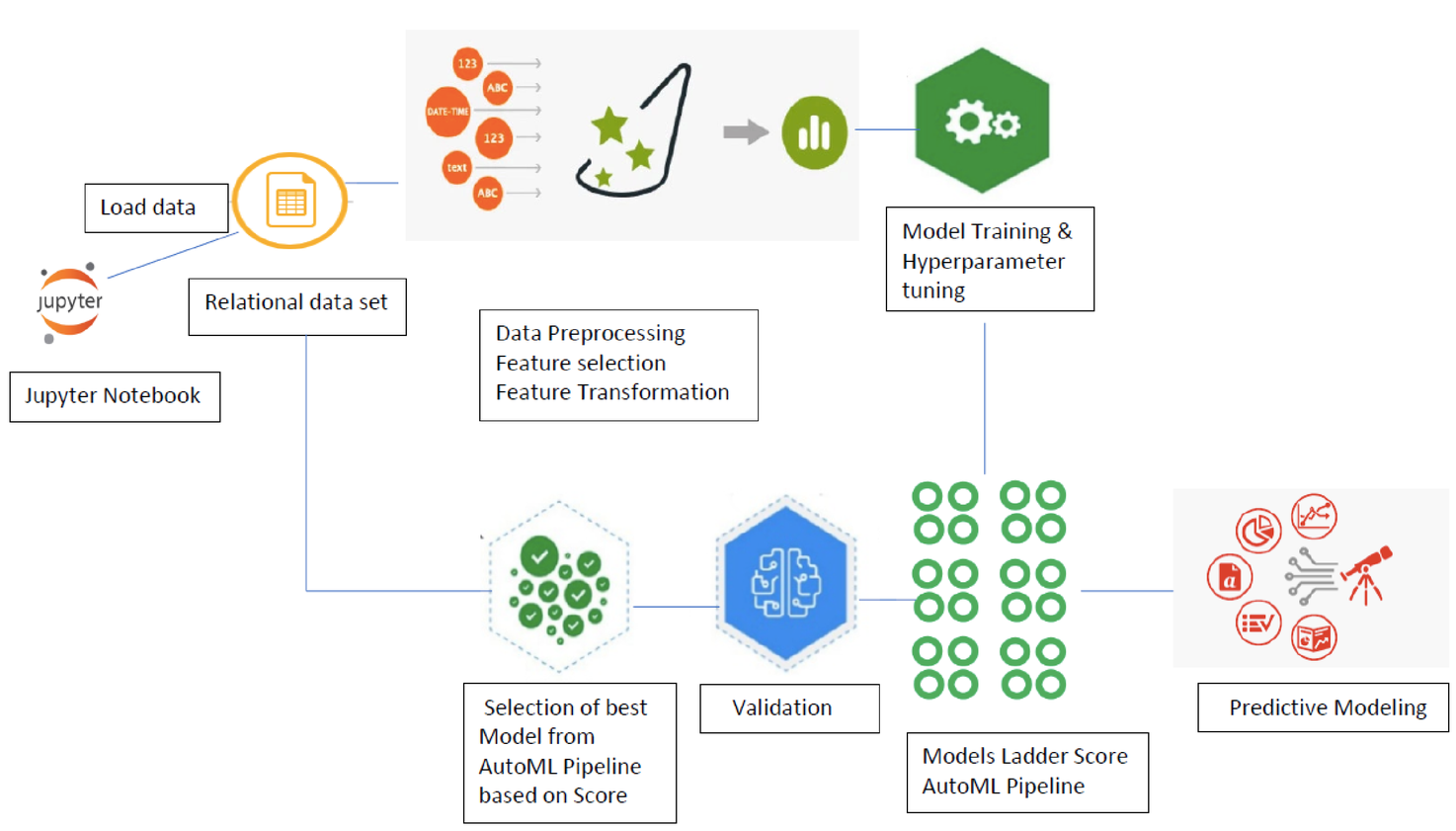
[1] Galhotra, Sainyam, et al., Automated feature enhancement for predictive modeling using external knowledge. In: International Conference on Data Mining Workshops (ICDMW). IEEE, 2019.

[2] Cheng, Weiwei, et al Automated feature generation from structured knowledge." Proceedings of the 20th ACM international conference on Information and knowledge management. 2011.

[3] Kanter, James Max, and Kalyan Veeramachaneni. Deep feature synthesis: Towards automating data science endeavors." 2015 IEEE international conference on data science and advanced analytics (DSAA). IEEE, 2015.

Automatic Machine Learning Architecture

Automated machine learning (AutoML) is the process of automating machine learning tasks.stages of AutoML are (**A**)Data Preparation,(**B**)Feature Engineering,(**C**)Model Training Hyper-parameter tuning, (**D**)Model Selection,(**E**)Validation and Score.



Our contribution

We Proposed Automatic Feature Engineering Framework, the framework is equipped with Automatic Knowledge Graph, **AutoKG** is generated from knowledge source wikipedia.

The AutoFE framework enables the addition of relevant new features using knowledge.