Post-mortem and modeling rationale.

Nalinda Kulathunga & Ryan Kennedy (University of Houston)

UNIVERSITY of HOUSTON

Problem statement / Aim

Introduction : People are quite interested in giving predictions on results of important world and local events such as elections, currency exchange rates, gold prices etc. They use different news sources, their own intuition or many other techniques to better predict the events. We are interested in using and fine tuning machine learning models to predict the reliability of the reasons people provide to support certain predictions : rationale

Aim : To train a model to learn from rationale features and predict the reliability of the rationale.

Data Set

Features : In the row data set, there are information about the question asked, the time frame of the answer, exact text of rationale and more. First we have used the Linguistic inquiry and word count (LIWC - 2015) package [1] to find possible feature space. Further more, time between creation of the question and resolving, and the presence of web information in text are also used as features.

Labels : Brier score (A score calculated based on the probability given by user and the correctness of the prediction). As a start, only two labeled classes were used. (good : <= 0.4, bad : > 0.4).

Tools

Language : Python2.7.

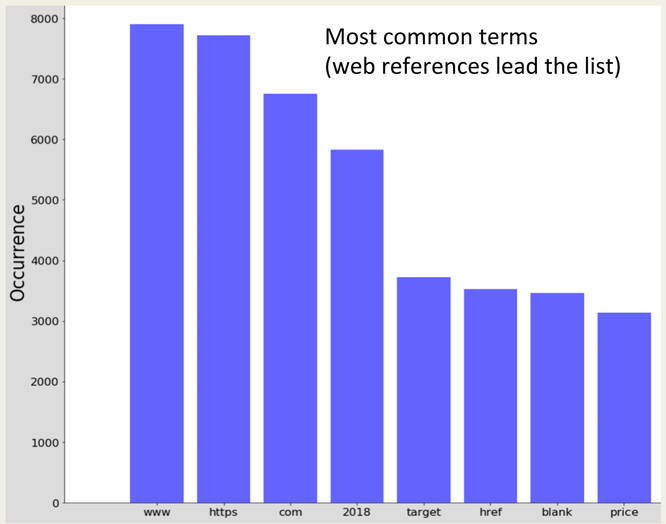
Stores : pandas data frames, numpy arrays.

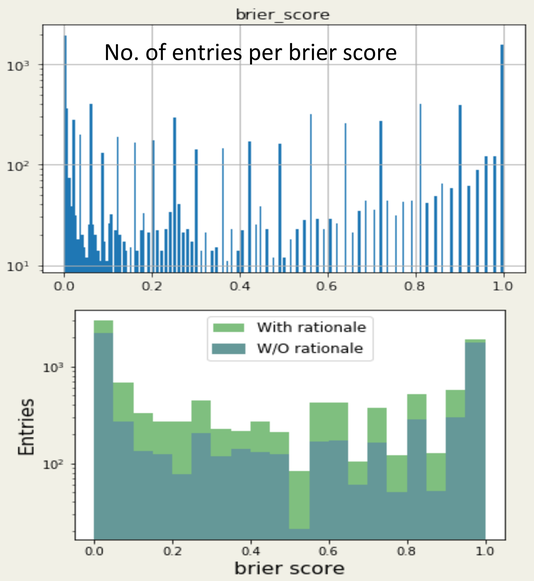
Models : From scikit learn package.

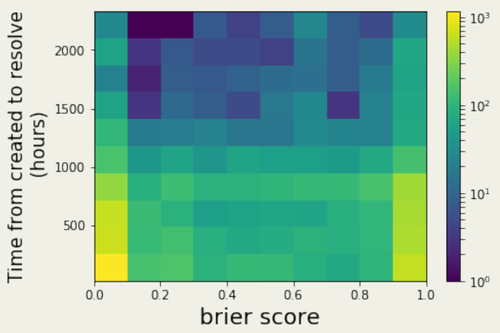
Other : NLTK [2], LIWC [1]

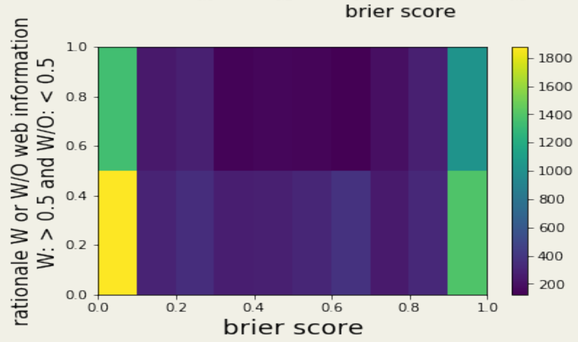
Feature Exploration

Before feeding the models, some already available features in the data set are examined for presence of possible classes in brier score. Effect of the brier score is tested with different approaches such as presence of a rationale, presence of web reference and time taken to resolve.





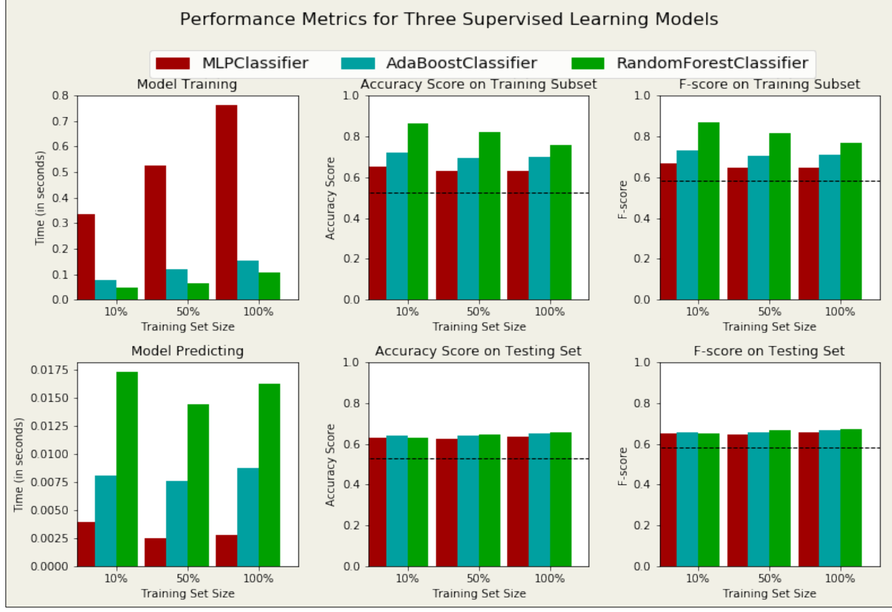




There seems to be a considerable presence of web references in the rationale. So we added the presence of web reference in rationale to the feature list. Even though there are no significant classes associated with other features listed above, they were added to the feature list in case of presence of any hidden classes.

Model Performance

Features highly correlated with brier score and presence of web information, time taken to resolve are used to feed the models. Best performed model details are as follows:



- **From this preliminary test on data, we have achieved an accuracy greater than a naive guess (best : Random forest). Tests on Improving the accuracy are in progress..!!**
- **After achieving a reasonable accuracy, next step would be modeling subjective quality.**
- **For that, human coders are evaluating the quality of rationale in terms of relevance, type of information, etc.**
- **Finally, these evaluations will be modeled using a similar features, methods used in analysis with brier score.**

Future Directions/Improvements

- Improving the feature selection and Fine tuning the models.
- Increasing the number of classed in the brier score.
- Ultimately, treat the brier score as it is (continuous brier score).
- Try to work with models help predicting continues labeling (eg: Support Vector Regression - SVR).
- Implement the steps needed for the subjective quality part of the analysis.

References :

[1]. Linguistic Inquiry and Word Count (LIWC), https://liwc.wpengine.com

[2]. Natural Language Tool-Kit (NLTK), https://www.nltk.org