Kiva

•••

Fast Money Helper

Microlending

- Peer to peer, small loans to impoverished borrowers
- Lack literacy, collateral, steady employment, or a verifiable income history
- Loans designed to support entrepreneurship and alleviate poverty
- Issued by individuals rather than banks

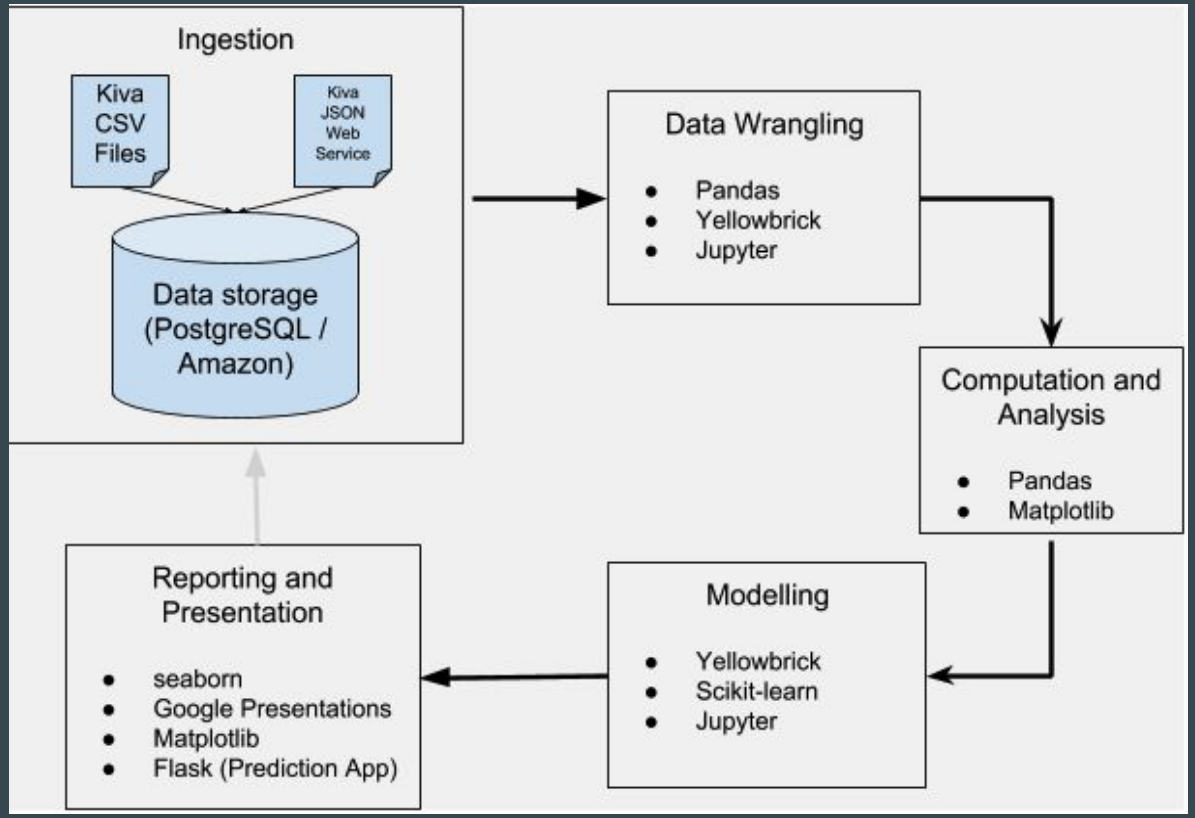
What is Kiva?

- Kiva's goal is to "*celebrate and support people looking to create a better future for themselves, their families and their communities. Kiva operates in the microfinance space through the Internet and connects borrowers and lenders using financial institutions in the borrower's country*"
- Available data
  - Loans
  - Lenders
  - Loan to Lender
  - Partners (through API)

Team Goal:

Our original goal was to explore determinants of loan repayment. However, after exploratory analysis, we decided to identify factors that determine the faster funding of the loans.

Data Science Pipeline / Architecture



Ingestion (Kiva API)

Login | Register | Forum | Kiva.org

BUILD.KIVA

Home Docs ▾ API My Apps

HELLO

We're Open!

Developers worldwide are helping us make it easy and transparent to lend to the working poor via microfinance and the Kiva API. This is the home for all the information and tools you need to join us!

Learn

Check out our developer docs or explore the API Reference.

Build

Create the application Kiva Lenders have always dreamed of.

Share

Let the world experience microlending through the applications you've built!

Ingestion (Kiva API... continued)

- Using a data snapshot, we had access to:
  - 1,419,607 Loans
  - 2,349,174 Lenders
  - A relationship table to link the two tables together

- Using the JSON python toolkit, we had access to:
  - 503 Partners (banks who connect the lenders to the borrowers)

Ingestion (Database)

- Using our Georgetown email address and AWS Educate, we were able get $75 of credit for AWS.

- We created a PostgreSQL database to store our Kiva data very early in the process

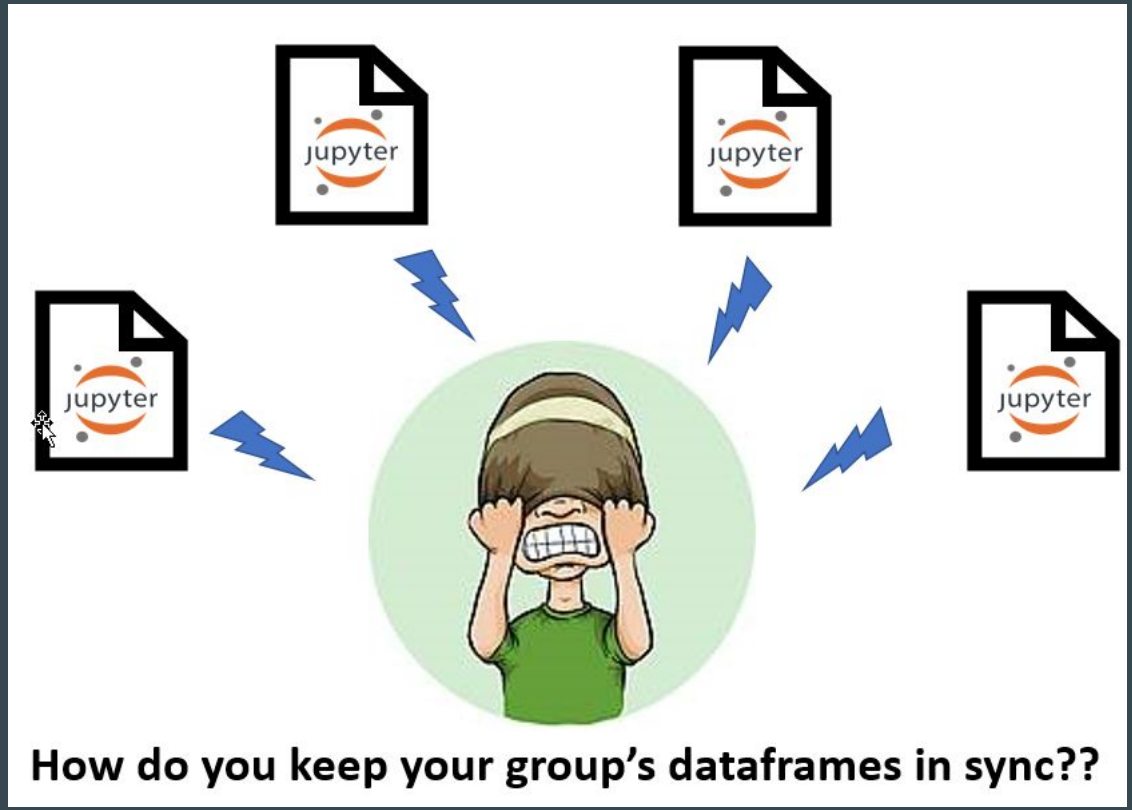
“Don’t be that team that uploads your AWS credentials to Github!”

- Allen Leis

Solution:

- Create a python file that contains your credentials as a variable
- Add that file to .gitignore
- Import the variable in Jupyter or other classes
- Provide an example file (e.g. dbconfig.py.example)

Wrangling everyone's ideas and code...

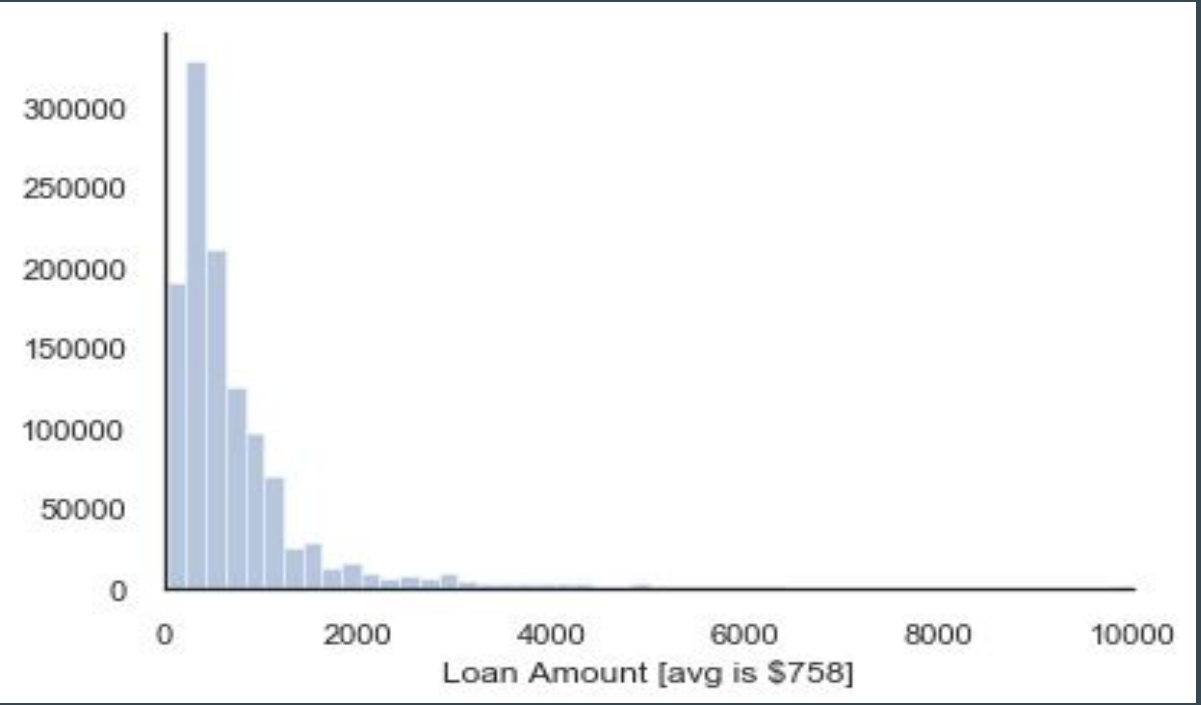


KivaDataLoader.get_clean_dataframe()

- Our requirement:
  - One dataset
  - Maintainable by anyone using Python on Github

- Solution
  - Combine the SQL and Python transformations into one callable function.
  - Returns a dataframe with all required columns and data

Exploratory Analysis



Sample Characteristics

Over one million loans [instances]

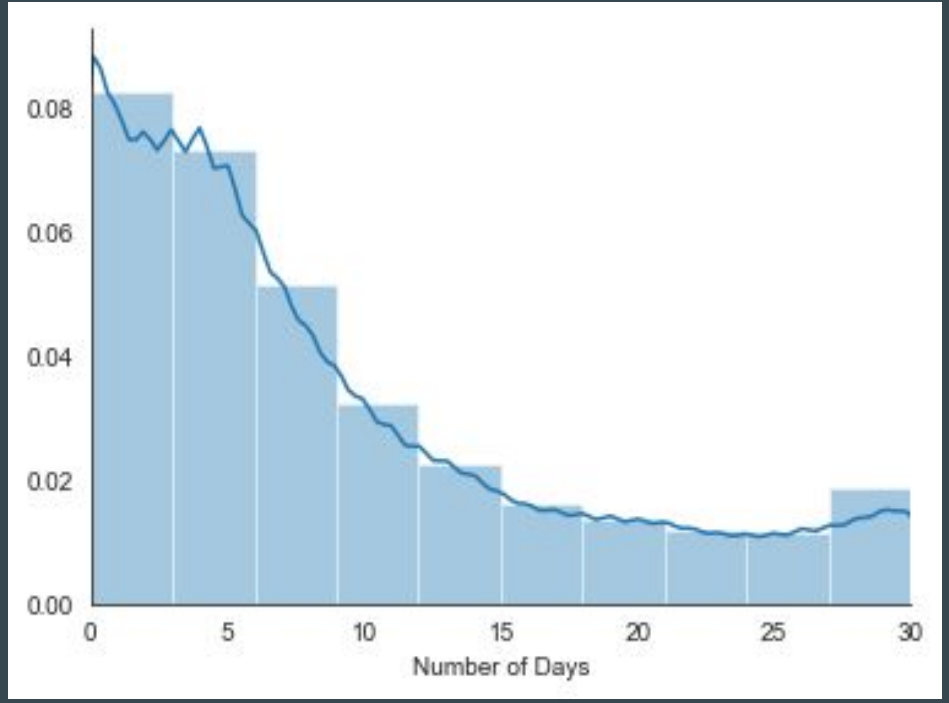
58 Features

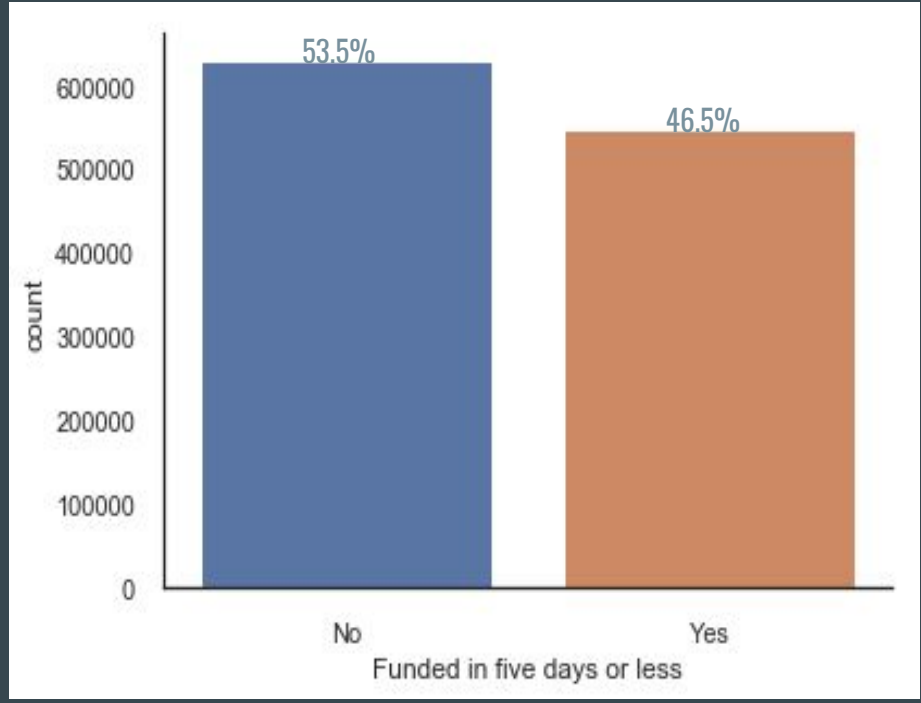
96 Countries

15 Sectors

12 years

Exploratory Analysis (continued)





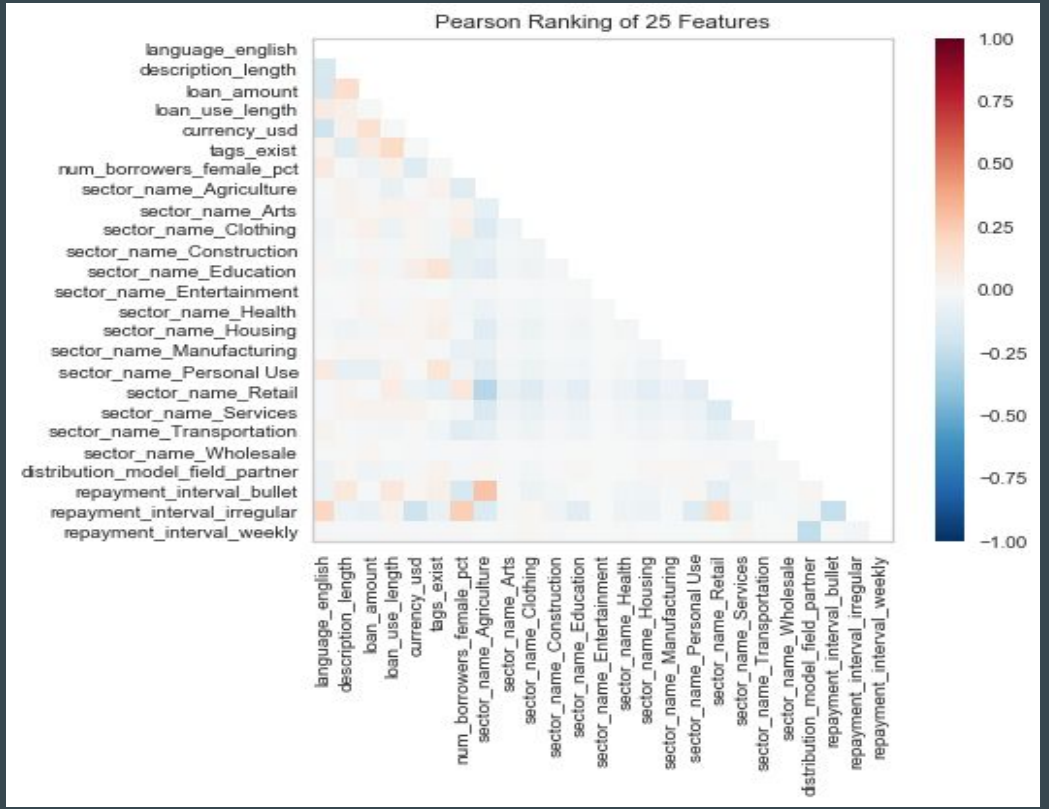
Posted to raised

Target Variable

Feature Selection

Features

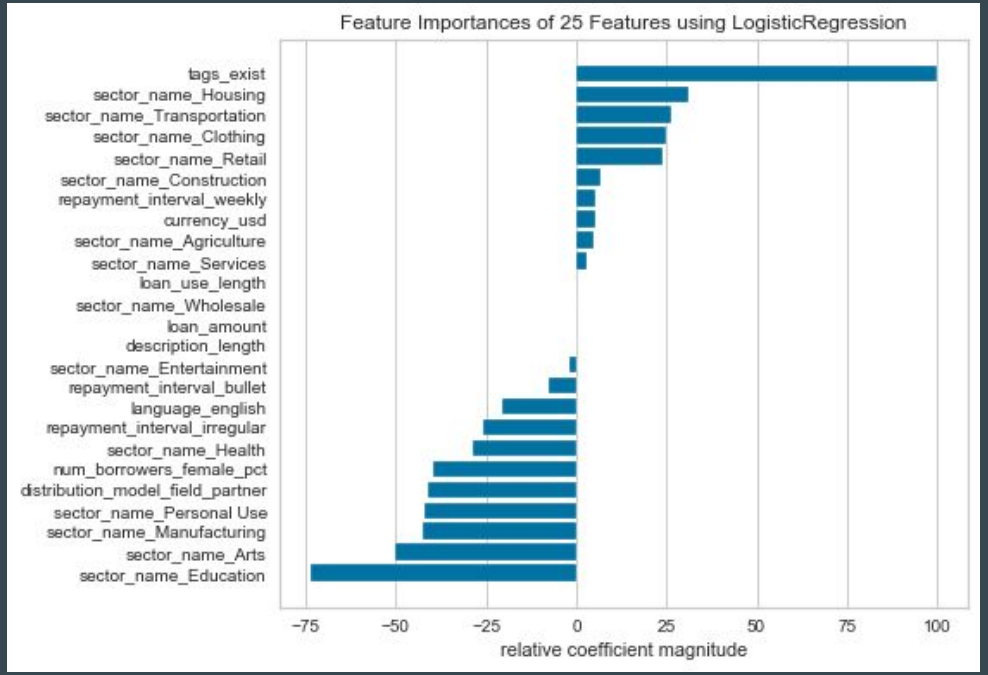
- Language
- Description
- Loan use
- Currency
- #Hashtag
- Gender
- Sector
- Loan type
- Repayment schedule

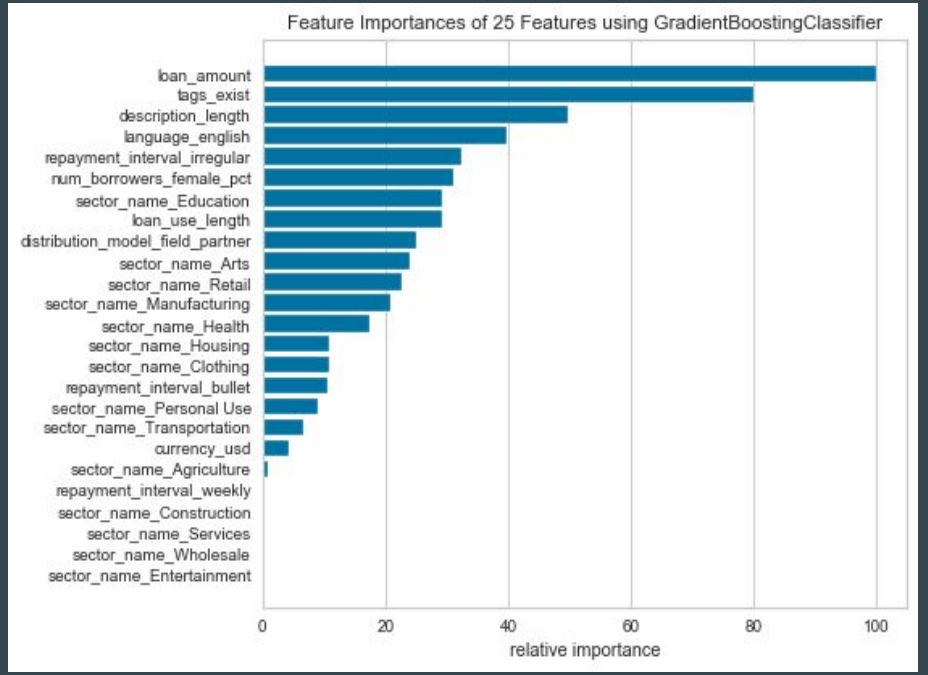


Model Selection

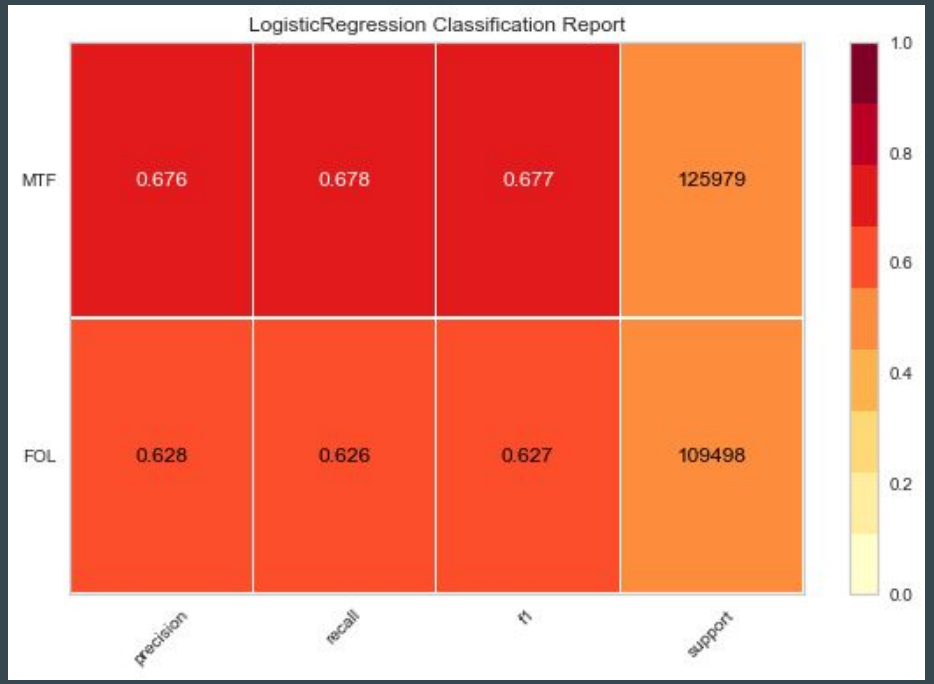
Classifier	Accuracy	Log Loss
Gradient Boosting	65.8%	0.62
Logistic Regression	64.7%	0.63
Random Forest	64.5%	0.67
Multi-layer Perceptron	63.2%	0.74
Gaussian NB	58.9%	1.05
Decision Tree	58.5%	14.28
KNeighbors	54.8%	4.71
Linear SVC	53.7%	

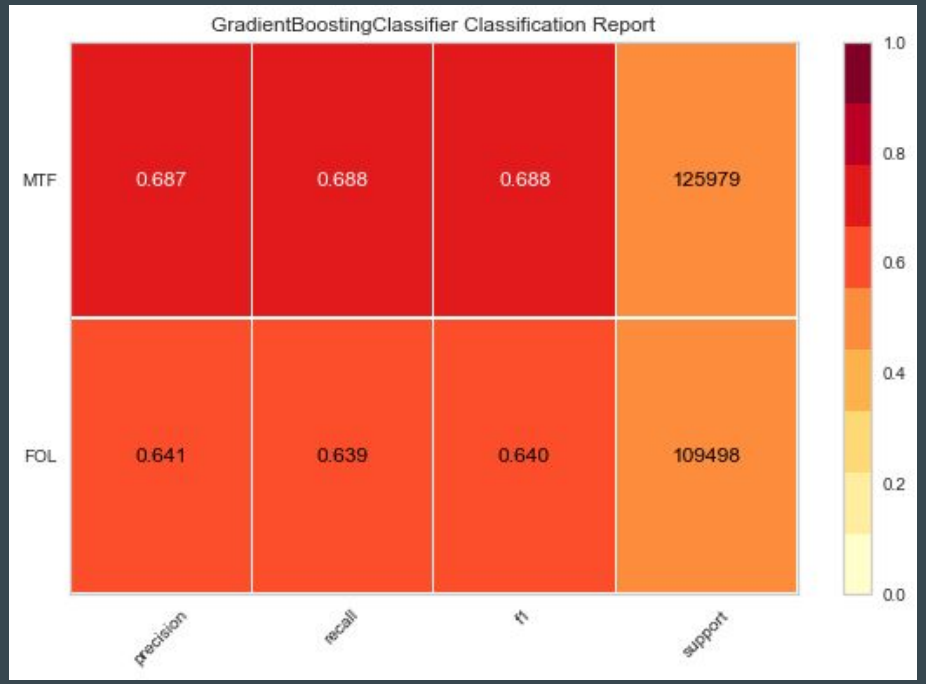
Model Selection (continued)





Model Selection (continued)

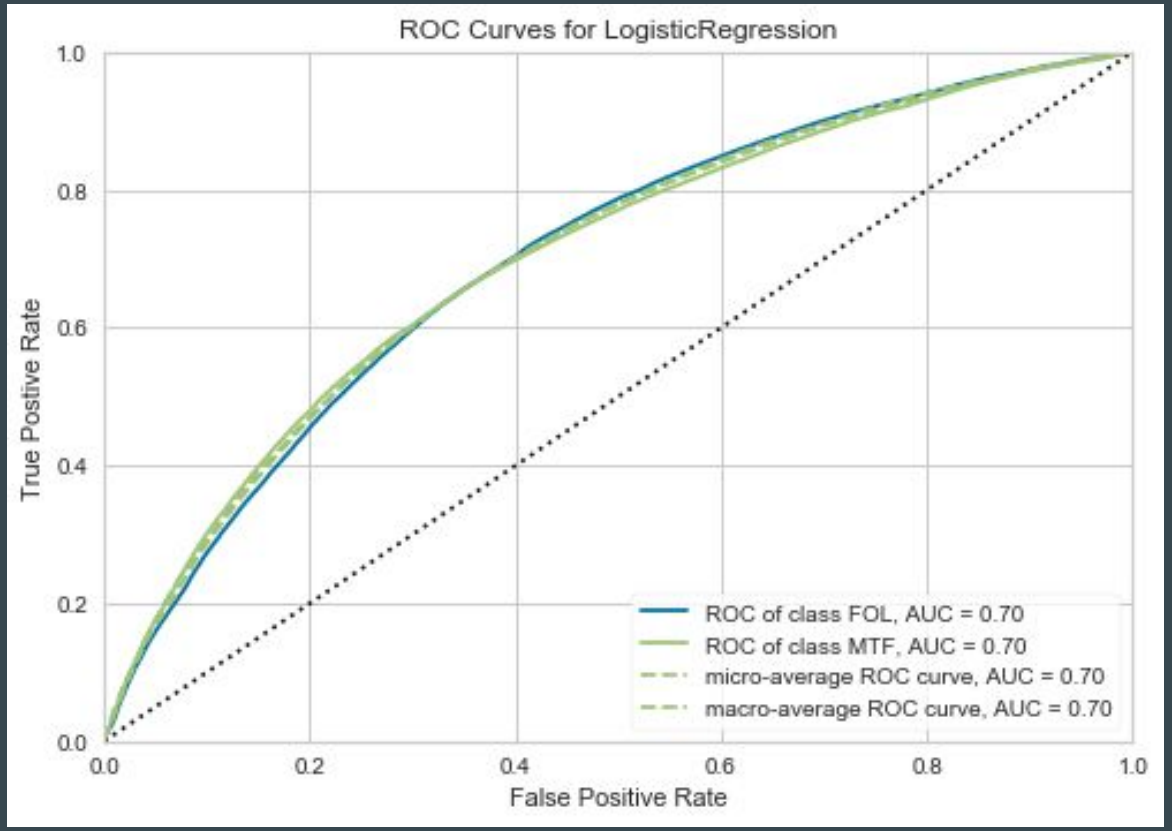




Runtime = 5.77s

Runtime = 101.49s

Modelling: Aiming for the Triple



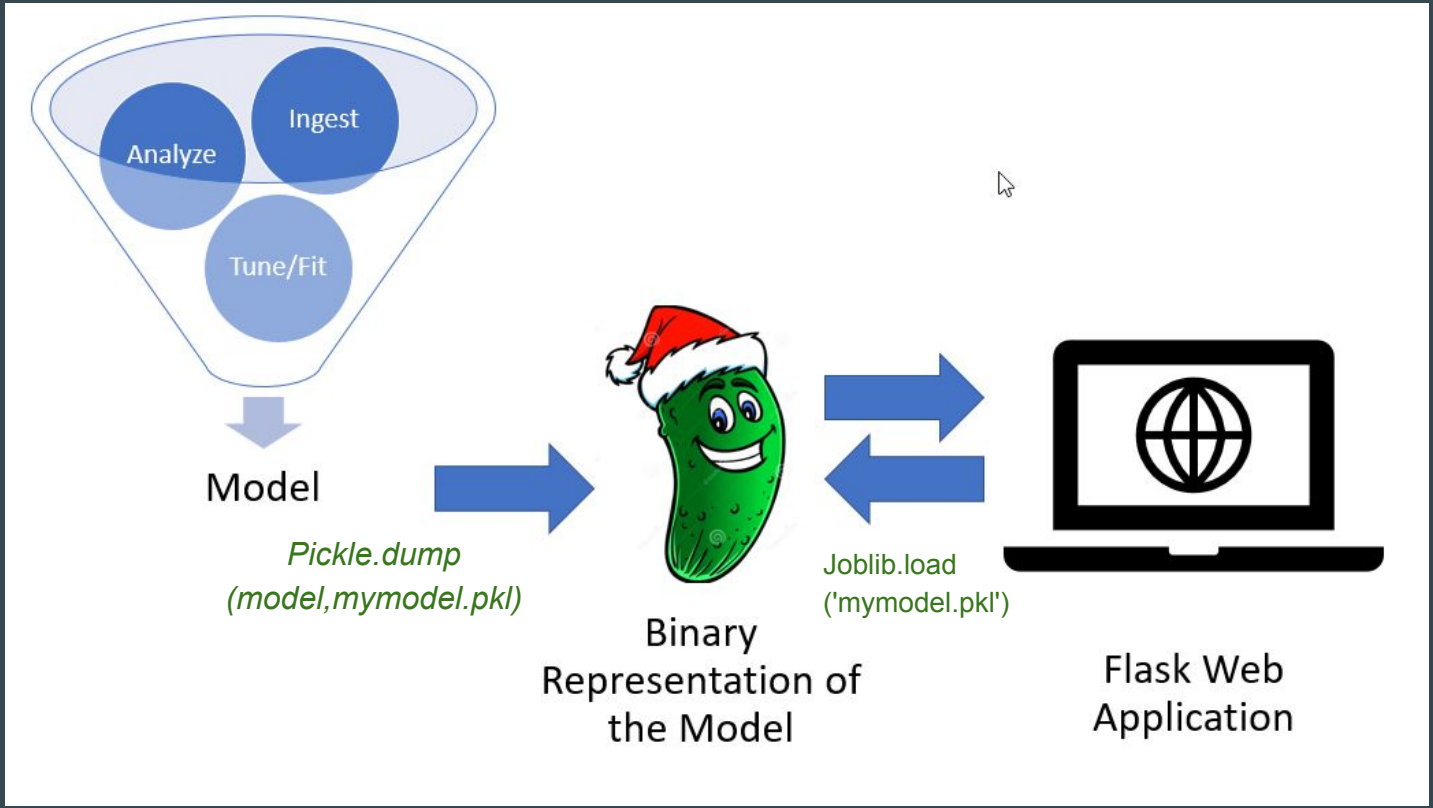
Model Tuning

Hyperparameter tuning
- GridSearchCV
- Regularization

Scalers
- StandardScaler
- Normalizer
- MinMaxScaler
- MaxAbsScaler
- RobustScaler

Additional features
- Loan_distance

The Flask Application Explained



https://www.123rf.com/photo_57291639_stock-vector-christmas-pickle-cartoon.html

Flask demo

http://kivateam.pythonanywhere.com/

Conclusion

- Hypothesis:
  - Initially focused on loan repayments/lender side - to predict success/failure of a loan.
  - Redefined project goal and built a model to predict time taken for a loan to get funded/borrower side.
- Team dealt with large data set - over 1.4 million incidences with more than 58 features - Kiva API.
- A lot of data wrangling and data transformation performed.
- Skewness of the target variable was the biggest challenge.
- Regression models didn't produce desired results therefore team resorted to classification models.
- Class imbalance was also a problem with classification models.
- Team worked with binary classification to overcome the problem of class imbalance.

Conclusion (continued)

- Results: model established the right relationship between the target variable (posted_to_raised days) and features incorporated in the model.
  - The size of a loan a borrower requested is negatively related to the time it takes for a loan to get funded.
  - The better the description of the loan purpose (description_length), the faster a loan gets funded.
  - The shorter the loan term, the faster the loan gets funded and etc.
- Data product: the data product built by the team enables borrowers to check if they can get funded in as little as 5 days or if it will take more days.
- Potential future work: borrower side to predict success/failure of a loan to improve loan repayment rates.

- Team members:
  - Tom O'Neil (Team Spokesperson)
  - Yamil Vargas
  - Nuredin Abdella
  - Won Choe