Multi-Class Text Classification with Kiva Loans Descriptions and Sectors

Ting Ting Liu

KIVA
Loans that change lives

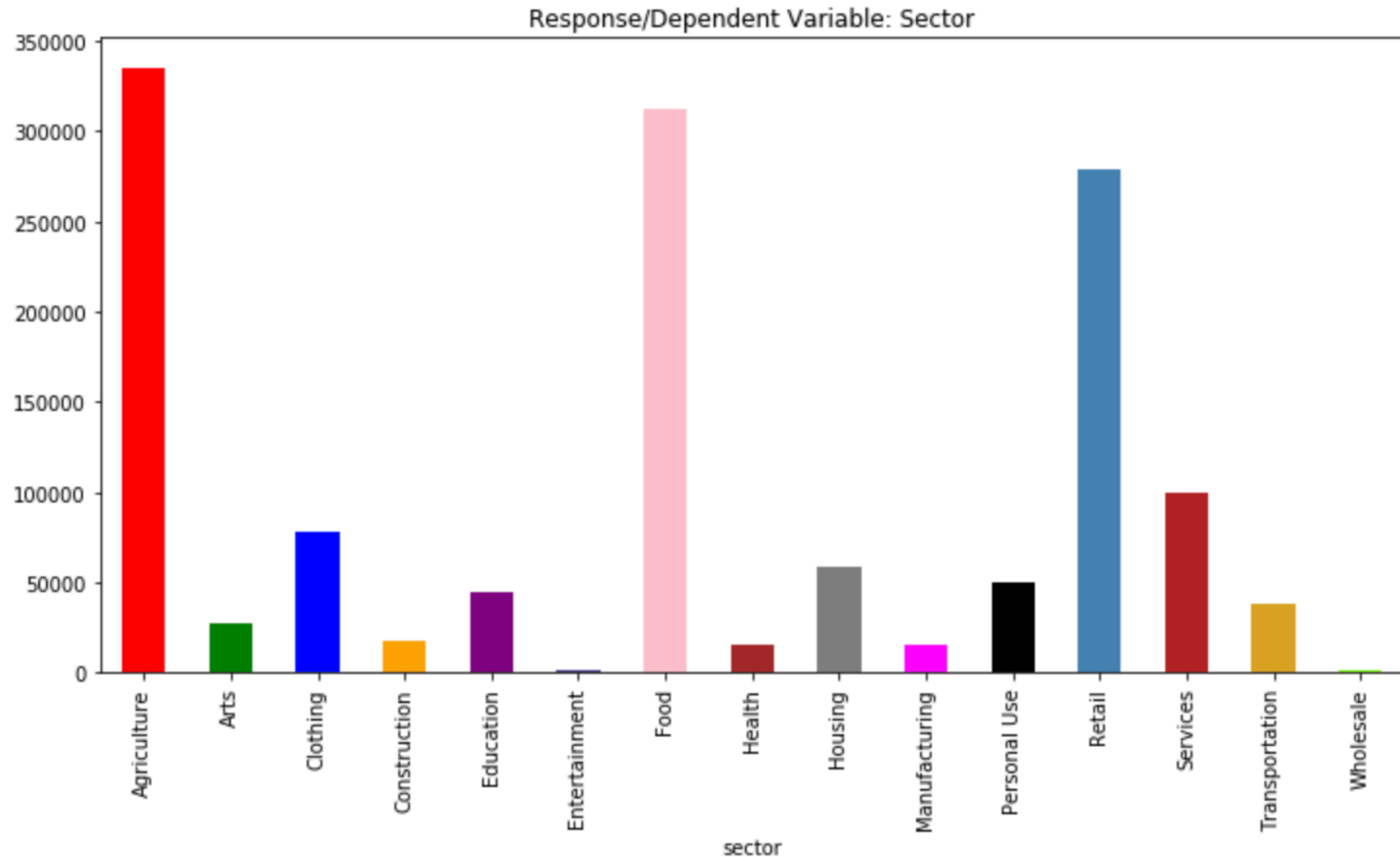# Data: Kiva.org Loan Listings

This dataset is from the Kiva API (https://build.kiva.org). More specifically, this is Kiva's Loans Data Snapshot, which contains all the loan listings that Kiva has posted from 2006 till January 2018. The dataset contains 1.4M rows and 34 columns. Each listing contains information that shows the key details of a loan, including name, loan amount, loan description, sector, etc.

```
loans.csv:
|-- posted_time: the datetime the loan was posted on the Kiva.org website (datetime)
|-- original_description: the description of the loan in the language of the country the
borrower is from (string)
|-- translated_description: the description of the loan translated to English if the
original language is not English (string)
|-- sector: the specific category of the loan (categorical)
```

kiva

Summarized Data: Imbalanced Classes

Response/Dependent Variable: Sector

Text Representation TF-IDF

'Arts':
. Most correlated unigrams:
    . sequins
    . embroidery
. Most correlated bigrams:
    . sequins pearls
    . embroidery business
. Most correlated trigrams:
    . sequins pearls various
    . pearls various colors

'Food':
. Most correlated unigrams:
    . ingredients
    . fish
. Most correlated bigrams:
    . buy ingredients
    . vending business
. Most correlated trigrams:
    . nwtf buy ingredients
    . vending business philippines

kiva

Text Representation TF-IDF

```
# 'Construction':                           # 'Retail':
. Most correlated unigrams:                 . Most correlated unigrams:
        . centering                                 . general
        . timber                                    . canned
. Most correlated bigrams:                  . Most correlated bigrams:
        . making bricks                             . general store
        . sells timber                              . store philippines
. Most correlated trigrams:                 . Most correlated trigrams:
        . selling construction materials            . store philippines requested
        . business selling construction             . general store philippines
```

kiva

Fit Model: Multi-Class Text Classification Naive Bayes Classifier

Models	Train Accuracy	Test Accuracy
Used TF-IDF Vectorizer as features	79.21%	77.36%
Used Count Vectorizer as features	78.98%	77.26%

kiva