

All files mentioned in this file are uploaded into the *github* repository.

The \LaTeX code involved in the generation of this document was aided by the example code provided in the links that Dr. Nelson sent out on January 17, 2016[1].

This document was compiled on www.sharelatex.com

Problem 1

The rss feed that was used in this assignment was obtained from Sports World News 4/12/16 at 4:15 PM EST, via the following link <http://www.sportsworldnews.com/rss/>. Once this feed was obtained, the 100 entries were then classified according to the following categories as seen in Table 1.

| Number | Category |
|--------|--------------------------|
| 1 | Wrestling |
| 2 | National Hockey League |
| 3 | National Football League |
| 4 | Boxing |
| 5 | Other |
| 6 | MLS |
| 7 | Major League Baseball |

Table 1: Categories and their description

Problem 2

A program called `makeTrainingData.py` was created to aid in assigning each feed a category. The program reads a list of categories and the name of an rss feed link or file. It then displays the description in the entry and prompts the user for a number corresponding to a category as seen in Figure 1. Afterwards, the user is prompted for the name of a file to save the data, and the program saves the data in the format `<guid> | <category>`.

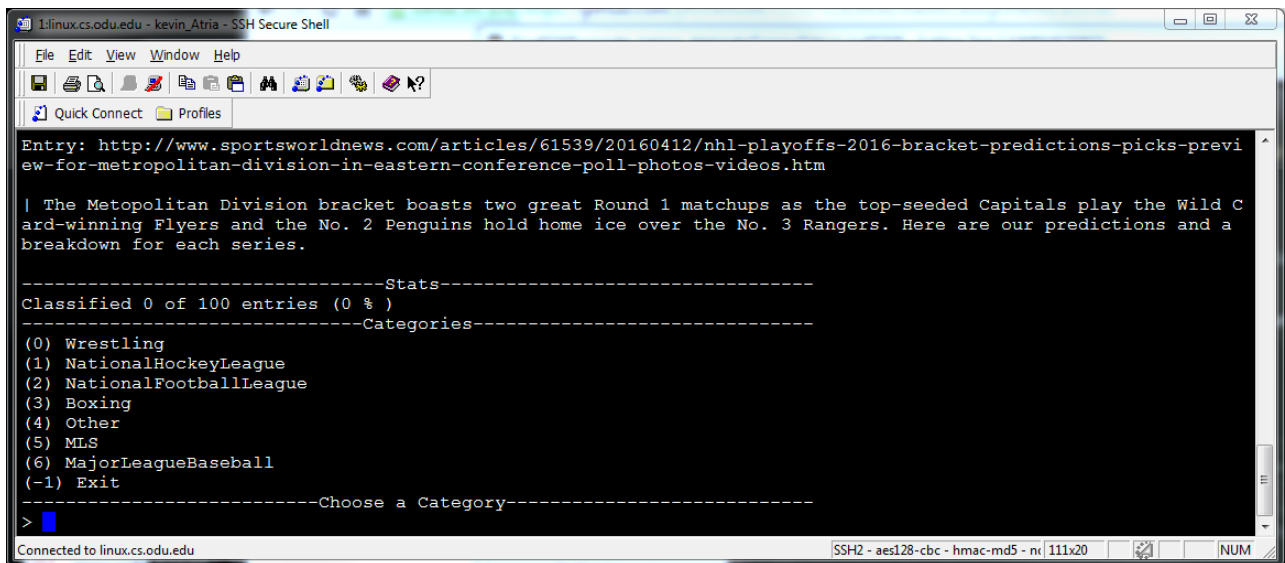


Figure 1: Screen-shot of makeTrainingData

Once all 100 entries were classified, a program called trainClassifier.py was written. This program utilizes a sqlite3 database. This database is also accessed by fisherclassifier[2]. First, the title, guid(link), description and actual category, are parsed and inserted into a table called feeds which is located within the database. This table also has an integer called isTrainingData for which 0 means that is is not training data and 1 indicating that is is training data. A dictionary is also created which maps the guid to the feedelement created by the feedparser. Once the information from all 100 feed elements are extracted, and loaded into the database, a file called partialTrainingData.txt is read by the program. Each line of the file is in the format <guid> | <category>. The program matches the guid with the description in each feed acquired from the database and uses the description to train the fisher classifier[2] and updates the corresponding column isTrainingData to a 1. classifier. The database code in the fisher classifier was rewritten to use sqlite3.

Listing 1: Training the fisher classifier

```
1 def train_classifier(settings, trainingData):
2
3     counter = 0 # Used for status information.
4     size = len(trainingData)
5     database = FeedDatabase(settings['database'])
6
7     # Create the classifier
8     classifier = fisherclassifier(getwords)
9
10    # Set the name of the database
11    classifier.setdb(settings['database'])
12    for key in trainingData.keys():
13        sys.stderr.write('...Training
14                          ({0}/{1})...\n'.format(counter, size))
15        classifier.train(trainingData[key]['description'],
16                        trainingData[key]['category'])
17        counter += 1
18    sys.stderr.write('...Finished Training Classifier\n')
```

After the fisher classifier is trained, a program called classifyFeed.py pulls the training data from the database. It then pulls all the feed entries in which the value is TrainingData is equal to zero and then classify those entries based on the content in the description. After the remaining unclassified entries are classified, the guid, and predicted category are printed to the standard output. For this assignment, the standard output is piped into a file called resul2.txt. Then a script called analyzePredictions.py, pulls the actual categories for each feed from the database, compares them with the predicted categories in the file result.txt and computes the fisher probability. The title, predicted category, actual category and fisher probability are then displayed to the standard output. Table 1 shows the results from the analysis.

| Title | Predicted Category | Actual Category | Fisher Probability |
|---|----------------------------|----------------------------|--------------------|
| J.J. Watt Girlfriend: Texans DE Spotted On Date With Kate Hudson [VIDEO] | National Basketball League | National Football League | 0.310612856966 |
| Air Jordan Release Dates & Price: AJ 17+ Retro Copper and AJ 4 Retro Pinnacle Coming Out [Photos] | National Football League | Other | 0.0949490510526 |
| NBA Trades: Los Angeles Lakers Almost Dealt Kobe Bryant to Detroit Pistons in 2007 | National Basketball League | National Basketball League | 0.258121846233 |

| | | | |
|---|----------------------------|----------------------------|-----------------|
| Los Angeles Lakers Rumors: Iggy Azalea Threatens To Cut Off Nick Young's Penis If He Cheats On Her Again [VIDEO] | National Basketball League | National Basketball League | 0.553163568105 |
| Bill Simmons HBO News: Game Of Thrones Season 6 Gives Sports Guy A New Gig [VIDEO] | Wrestling | Other | 0.088246841396 |
| Barcelona vs Atletico Madrid: Champions League 2016 Quarterfinals Live Stream, Where to Watch Online, Rosters | Major League Baseball | MLS | 0.0595697546653 |
| WWE Rumors: Next Opponent For Dean Ambrose Possibly Revealed [VIDEO] | Wrestling | Wrestling | 0.706240890025 |
| Nike Release Dates & Price: Kobe 11 'Lower Merion Aces' Pays Tribute to Bryants Alma Mater, Check Photos and Details Here | National Basketball League | National Basketball League | 0.691119394578 |
| WWE 'SmackDown' Spoilers 4/7/16: Match Results For 'WrestleMania 32' Fall-out Edition As Cesaro Returns [POLL, VIDEO] | Wrestling | Wrestling | 0.323058580156 |
| REDS vs CUBS Watch Live Stream Online & Listen: Opening Day For Wrigley Field, Live Coverage From Chicago | Major League Baseball | Major League Baseball | 0.949377763656 |
| The Undertaker WWE Return Rumors 2016: Another Sign Of 'Deadman's' Career Being Over Surfaces [VIDEO] | Wrestling | Wrestling | 0.487022043892 |
| WWE 'RAW' Spoilers & Rumors: AJ Styles, Roman Reigns Ready For Battle [POLL, VIDEO] | Wrestling | Wrestling | 0.356032010374 |
| Rob Gronkowski Rumors: Patriots Star's Amazing Offseason Includes Photo Shoots With Beautiful Women [PHOTOS, VIDEO] | National Football League | National Football League | 0.755665853992 |
| NFL Trade Rumors 2016: Dallas Cowboys Eyeing to Deal for Carson Wentz? | National Football League | National Football League | 0.831531144167 |
| Aaron Rodgers, Olivia Munn Relationship Update: QB Lashes Back At Presidential Candidate John Kasich [VIDEO] | National Football League | National Football League | 0.62391633611 |
| Renee Young, Dean Ambrose Relationship Rumors: 'Lunatic Fringe' Better Not Appear On 'Total Divas' [VIDEO] | Wrestling | Wrestling | 0.846363747872 |

| | | | |
|--|----------------------------|----------------------------|----------------|
| Chicago Bulls Rumors 2016: Pau Gasol, Jimmy Butler Leaving This Summer? | Major League Baseball | National Basketball League | 0.27647413967 |
| New York Knicks News & Rumors: Phil Jackson to Leave? Hall of Famer Eyeing Return to Los Angeles Lakers? | National Hockey League | National Basketball League | 0.12807500029 |
| Sting WWE Retirement 2016 Rumors: Hall Of Famer Opens Up About Calling It A Career [VIDEO] | National Football League | Wrestling | 0.348206373073 |
| Aaron Rodgers, Olivia Munn Relationship Update: Actress Has Her Own Critics, Just Like QB Boyfriend [PHOTO, VIDEO] | National Basketball League | National Football League | 0.356949995972 |
| Golden State Warriors Record 2016: Michael Jordan Gave Blessing to Break 72-10 Card, Says Draymond Green | National Basketball League | National Basketball League | 0.303237792126 |
| Brie Bella Retirement WWE Rumors: Daniel Bryan's Wife May Not Have Had Her Last Match [VIDEO] | Wrestling | Wrestling | 0.513440510395 |
| Canelo Alvarez vs Amir Khan Predictions: British Superstar Could Pull Off an Upset, Says Roy Jones Jr. | Boxing | Boxing | 0.697330224212 |
| Transfer Rumors: Chelsea to Keep John Terry, Diego Costa? | National Football League | National Football League | 0.617262856773 |
| Brie Bella Retirement WWE Rumors: 'Total Divas' Key Reason For WWE Staying Mum About Superstar's Future [VIDEO] | Wrestling | Wrestling | 0.876076307374 |
| Kobe Bryant Wife: Lakers Star Uses Bling To Make Up For Past Mistakes With Vanessa Bryant | National Basketball League | National Basketball League | 0.467397246403 |
| TIGERS vs MARLINS Watch Live Stream Online & Listen: MLB Opening Day 2016, Live Coverage From Miami | Major League Baseball | Major League Baseball | 0.941471174738 |
| WWE 'Payback 2016' Rumors & Predictions: Potential Stipulations To AJ Styles vs Roman Reigns [VIDEO] | Wrestling | Wrestling | 0.207657485128 |
| WWE 'RAW' Spoilers & Rumors: Roman Reigns Era Begins With 'WrestleMania 32' Fallout [PHOTOS, POLL & VIDEO] | Wrestling | Wrestling | 0.379926617923 |

| | | | |
|--|----------------------------|----------------------------|-----------------|
| Colin Kaepernick Broncos Rumors: 49ers QB, Johnny Manziel Share One Common Plight [VIDEO] | National Football League | National Football League | 0.363125396532 |
| Greg Hardy News: Refusal To Discuss Domestic Violence Issue Based On Ray Rice Ban After Confession [VIDEO] | National Basketball League | National Football League | 0.583271727279 |
| Roman Reigns WWE News & Rumors: Company Not Doing Champion Any Favors By Making AJ Styles No. 1 Contender [POLL, VIDEO] | Wrestling | Wrestling | 0.712079829909 |
| Derek Fisher, Gloria Govan Rumors: Matt Barnes Still Wants To Fight Former Knicks Coach [VIDEO] | Wrestling | National Basketball League | 0.421703208553 |
| EPL Week 33 Schedule, Fixtures, Table & Live Stream: Leicester Continue Title Chase; Tottenham, Manchester United Collide in Crucial Match | National Basketball League | MLS | 0.0911292914534 |
| CM Punk WWE Return 2016 Rumors: UFC's Michael Bisping Reveals Why 'Best In The World' Should Have Stayed In Wrestling [VIDEO] | Wrestling | Wrestling | 0.528563606537 |
| Cody Rhodes Wife: Eden Stiles Appears Open To Wrestling In Ring Like Her Husband Stardust [VIDEO] | Wrestling | Wrestling | 0.508994321518 |
| Maria Sharapova Innocent Rumors: Melonium Inventor Says Drug Is Not PED, Ban Wrong [VIDEO] | National Hockey League | Other | 0.357778093453 |
| Floyd Mayweather vs Manny Pacquiao 2: Robert Garcia Says Rematch Will Happen If Pacman KOs Timothy Bradley | Wrestling | Boxing | 0.120440905181 |
| NFL Trade Rumors 2016: Los Angeles Rams to Trade Nick Foles to New York Jets or Denver Broncos? | National Basketball League | National Football League | 0.218609276649 |
| WWE 'Payback 2016' Rumors & Predictions: Potential Swerves To AJ Styles vs Roman Reigns [POLL, VIDEO] | Wrestling | Wrestling | 0.426511813761 |
| UFC Rumors 2016: Conor McGregor, Georges St-Pierre Could Face Off; Anderson Silva Wants Rematch with Michael Bisping | Wrestling | Wrestling | 0.659722056895 |

| | | | |
|--|--------------------------|----------------------------|----------------|
| Russell Wilson, Ciara Engagement Rumors: Singer Is Marrying Into Royalty [PHOTO, VIDEO] | Wrestling | National Football League | 0.260914518629 |
| DODGERS vs PADRES Watch FREE Live Stream Online: MLB Opening Day 2016 Live Coverage From San Diego | Major League Baseball | Major League Baseball | 0.841599973097 |
| Brie Bella Retirement WWE Rumors: 'Total Divas' Star Comments On Regrets Of Departing During New Women's Movement [VIDEO] | Wrestling | Wrestling | 0.878027940134 |
| Randy Orton Return 2016 WWE Rumors: 'Viper' Has 2 Opponents In Mind For When He Returns [VIDEO] | Wrestling | Wrestling | 0.57405658471 |
| Boxing News 2016: Canelo Alvarez Will Be Ready to Face Manny Pacquiao, Miguel Cotto [VIDEO] | Boxing | Boxing | 0.753799286872 |
| New York Knicks Rumors & News: Patrick Ewing Back to the Big Apple? Hall of Famer Wants to Coach in NY | National Football League | National Basketball League | 0.230202432222 |
| DODGERS vs GIANTS Watch Live Stream Online & Listen: Giants Home Opener, Live Coverage From San Francisco | Major League Baseball | Major League Baseball | 0.763208956982 |
| Lamar Odom, Khloe Kardashian Relationship Update: Reality Star Out Of Job After 'Kocktails With Khloe' Gets Canceled [VIDEO] | Wrestling | National Basketball League | 0.378533884798 |

Table 2: Predictions made from the fisher classifier

Problem 3

To analyze the performance of the classifier, a script called AS9P1.py which was graciously written by Gabriel Marquez, computes the precision, recall and F-Measure of the classifier. The formula that was used in the computation of the precision for a category is based on the formulas from wikipedia and is as follows[3]:

$$\text{Precision for a category} = \frac{\text{Number of correct predictions for a category}}{\text{Total Number of Predictions Made}}$$

To compute the recall the following formula was used:

$$\text{Recall for a category} = \frac{\text{Number of correct predictions for a category}}{\text{Actual Number of entries in the category}}$$

To compute the F-Measure for a category the following formula was used:

$$\text{F-Measure for a category} = \frac{2 \times \text{Precision for a category} \times \text{Recall for a category}}{\text{Precision for a category} + \text{Recall for a category}}$$

Table 3 shows the performance values computed for each of the categories.

| Category | Precision | Recall | F-Measure |
|--------------------------|-----------|--------|-----------|
| Wrestling | 0.32 | 0.72 | 0.44 |
| National Hockey League | 0.0 | 0.0 | 0.0 |
| National Football League | 0.1 | 0.38 | 0.15 |
| Boxing | 0.04 | 0.66 | 0.075 |
| Other | 0.0 | 0.0 | 0.0 |
| MLS | 0.0 | 0.0 | 0.0 |
| Major League Baseball | 0.08 | 0.66 | 0.14 |

Table 3: Performance values of the Fisher Classifier

Based on the measures provided in table, it is concluded that the training data is skewed in wrestling had the highest precision. This could be due to the fact that most of the training data was involved Wrestling as a category. It is suggested that if this assignment were to be continued or attempted again, that a training dataset should not show as much bias towards a single category. It is also hypothesized that perhaps performing ten fold cross validation may remove some of the bias that occurred in this assignment.

References

- [1] Mohammed Aturban. *Assignment 9*. 2013. URL: https://github.com/maturban/cs595-f13/blob/master/assignment9/latex/Mohamed_Aturban_Assignment_9.tex.
- [2] Toby Segaran. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O'Reilly, 2007.
- [3] unknown. *Precision and Recall*. URL: https://en.wikipedia.org/wiki/Precision_and_recall.