Text Sentimental Analysis on Facebook Posts

Samuel Beuret

Valentine Santarelli

Meryem Wehbe

Data Acquisition

1. **Sentimental Analysis Dataset:** 1.6 million collected labeled sentences with '0' for negative and '1' for positive. This dataset is used to train a 2 class classifier.

2. **Amazon Dataset:** 30k Amazon Product reviews from users with the rating given to the product. This dataset is used to train a multi-class classifier and Regressors.

3. **Facebook Dataset:** In order to study the reactions of the Facebook audience to a post , we needed to scrap

Here are parts of 3 functions used to recover the information from Facebook.

- **postReactions** (postId,filename)
- **lastPostsReactions**(page,filename,n)
- **getAllComment**s(postId,nb_comments_per_post,serie,fb)

Feature Extraction

Each word is represented by an integer and count the appearance in the document.

- **Removing Stopwords**: Using the English dictionary

- **Select K Best Model**: Selects the K best features in relation with the labels using the chi2 distribution.

```python
def compute_bag_of_words(text, stopwords , vocab=None):
    vectorizer = CountVectorizer(stop_words = stopwords,vocabulary=vocab)
    vectors = vectorizer.fit_transform(text)
    vocabulary = vectorizer.get_feature_names()
    return vectors, vocabulary

KBestModel = SelectKBest(chi2, k=1000).fit(bow, Y_sat)
indices = KBestModel.get_support(True)
bow_transformed = KBestModel.transform(bow)
```

Some of the **informative features** of our dataset: adorable, afraid, agree, attacked, awesome, awesomeness, awful, aww, badly, beautiful, celebrate, depressed, failed.

Classifiers and Regressors

```python
partial_fit_classifiers = {
    'SGD': SGDClassifier(alpha=0.0001, average=False, class_weight=None, epsilon=0.1,
        eta0=0.0, fit_intercept=True, l1_ratio=0,
        learning_rate='optimal', loss='hinge', n_iter=50, n_jobs=1,
        penalty='l1', power_t=0.5, random_state=None, shuffle=True,
        verbose=0, warm_start=False),
    'Perceptron': Perceptron(),
    'NB Multinomial': MultinomialNB(alpha=0.01),
    'Passive-Aggressive': PassiveAggressiveClassifier(C=1.0, n_iter=50, shuffle=True,
                                                      verbose=0, loss='hinge',
                                                      warm_start=False),
    'NB Bernoulli': BernoulliNB(alpha=0.01),
}
```

```python
partial_fit_Regressors = {
    'SGD Regressor':SGDRegressor(loss='squared_loss', penalty='l2', alpha=0.001, l1_ratio=0,
                        fit_intercept=True, n_iter=1000, shuffle=True, verbose=0, epsilon=0.01, random_state=No
                        learning_rate='invscaling', eta0=0.01, power_t=0.25, warm_start=False, average=False),
    'Passive-Aggressive Regressor' : PassiveAggressiveRegressor(),
}
```

Incremental Learning

The dataset is too large to train the classifiers at once **Memory Error** that is why we used incremental learning:

1. Divide Data into Batches

```
minibatch_size = 10000
batch = bow_transformed[start:start+minibatch_size]
```

2. Partially train the classifier on the batch

```
cls.partial_fit(X_train, Y_train, classes = y_all)
```

3. Cross-Validation

```
kf = KFold(n_splits = 10)
    for train_index,test_index in kf.split(X):
        X_train,X_test = X[train_index],X[test_index]
        Y_train,Y_test = Y[train_index],Y[test_index]
```

4. Accuracy Check:

Classifiers: `100*sklearn.metrics.accuracy_score(Y_test, train_pred)`

Regressors: `sklearn.metrics.mean_squared_error(Y_test, train_pred)`

Performance SAT
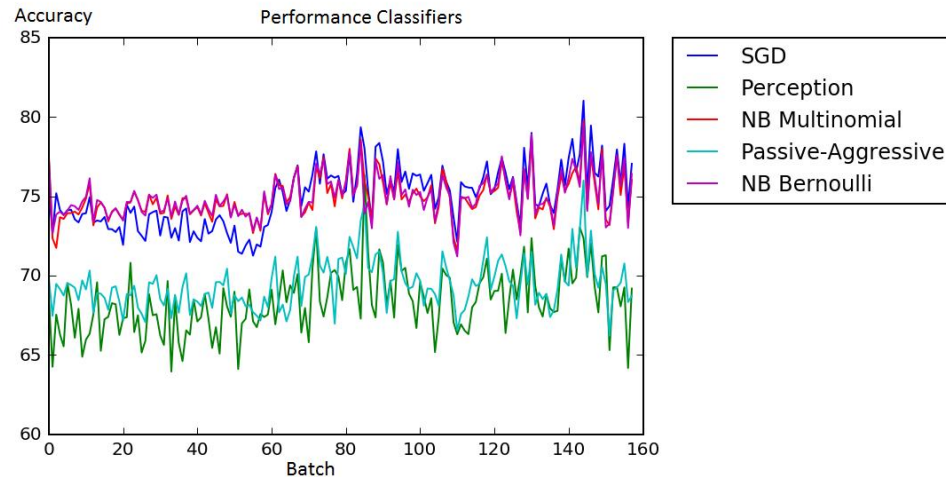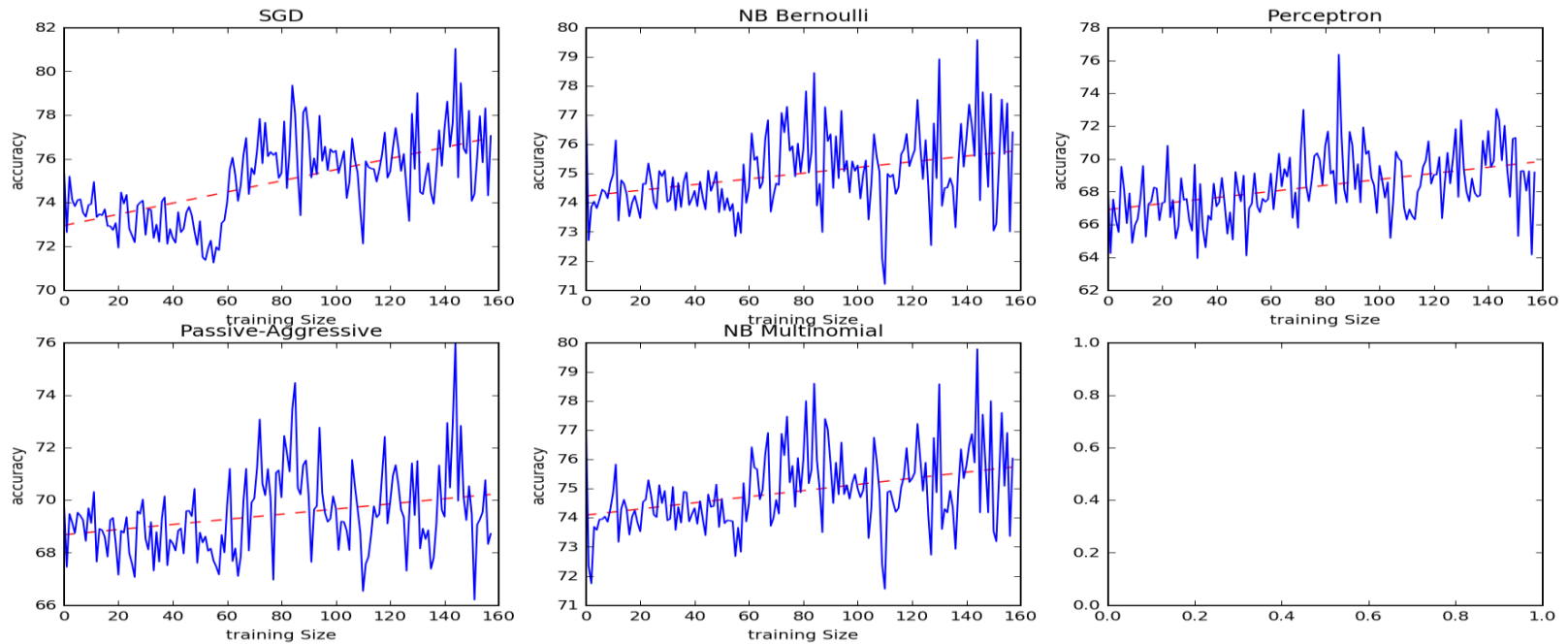
Performance Amazon
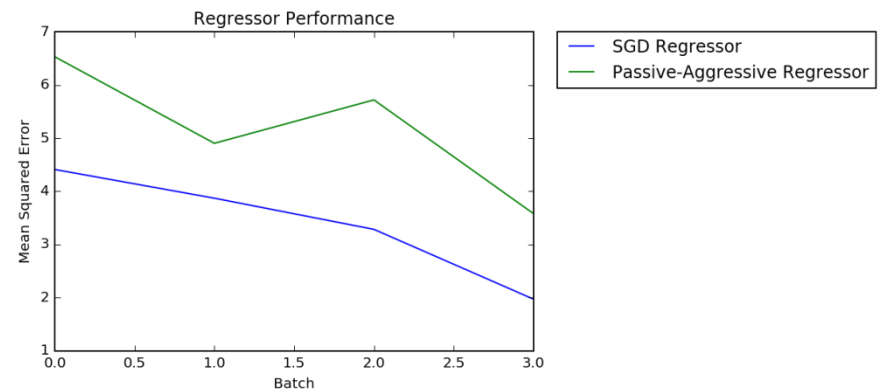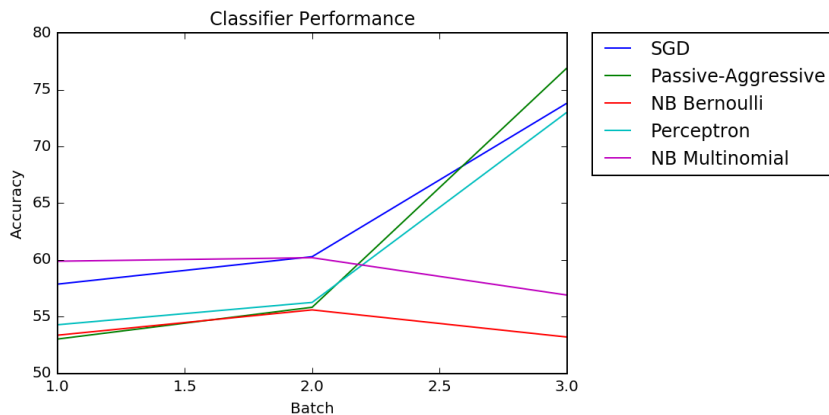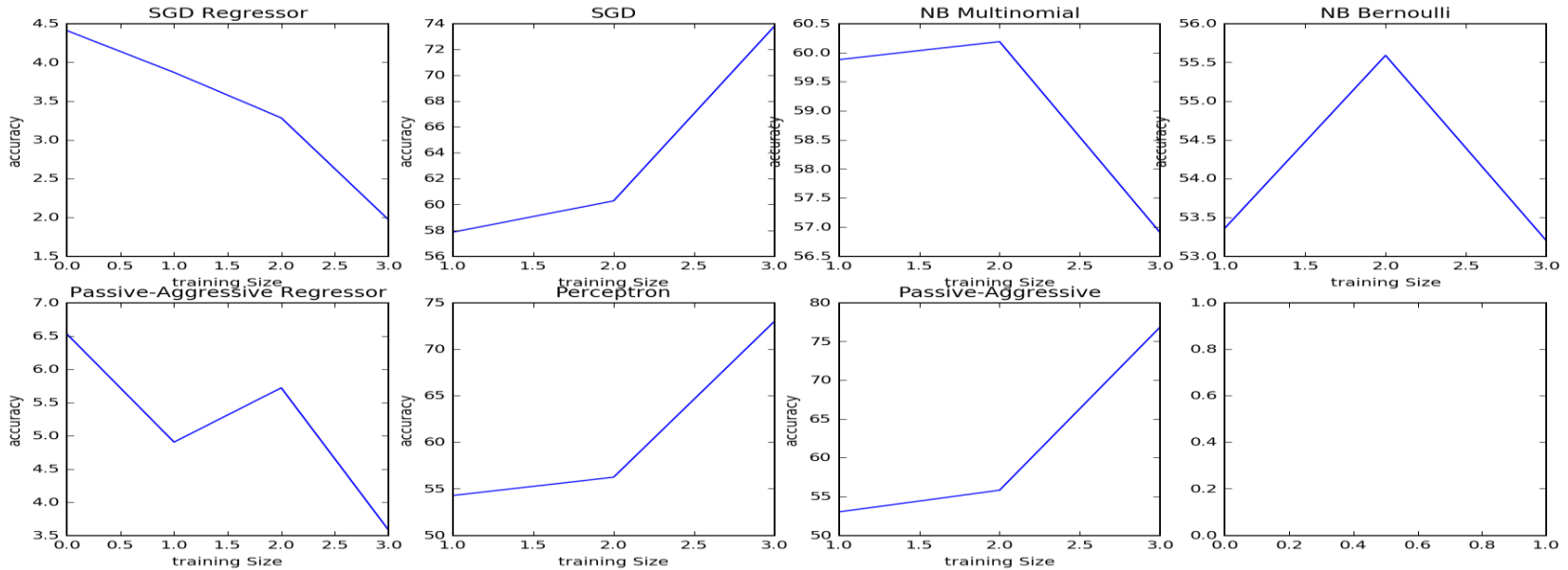
Data Analysis **Theme:** Terrorism

¶

CNN article about orlando shooting's victims suing twitter and facebook
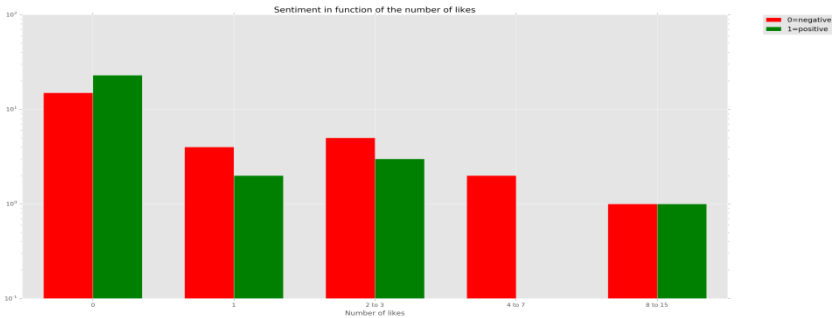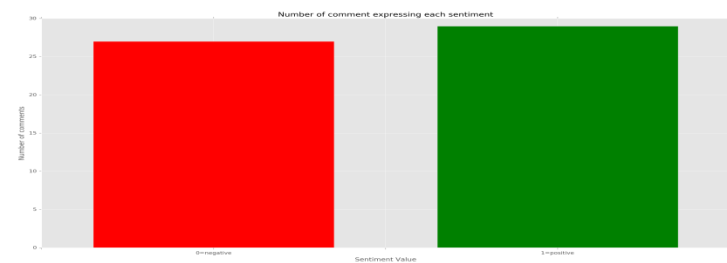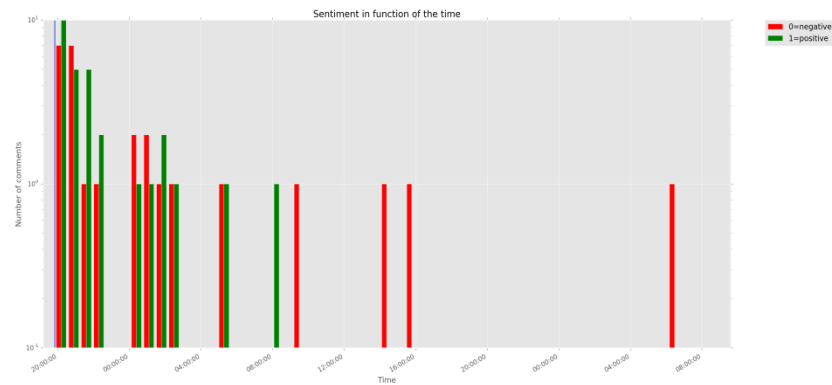
CNN article about John McCain saying that Obama is responsible for the Orlando shooting
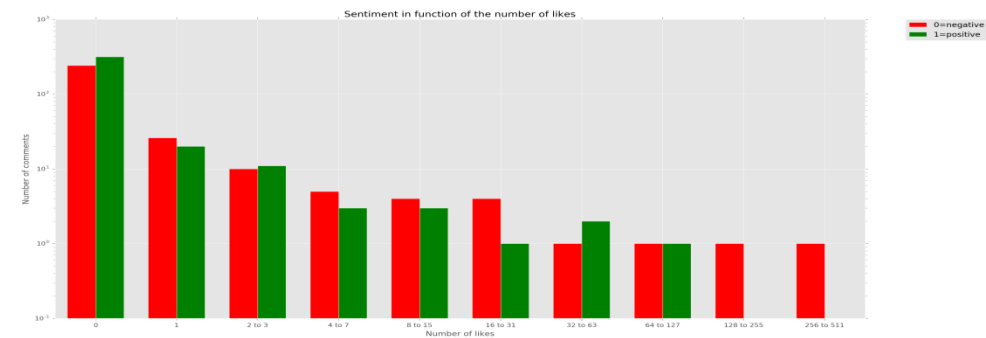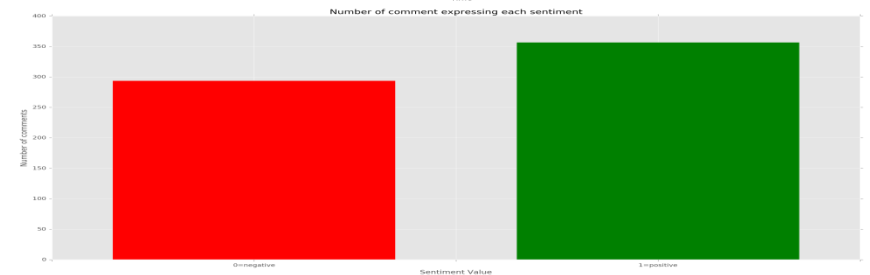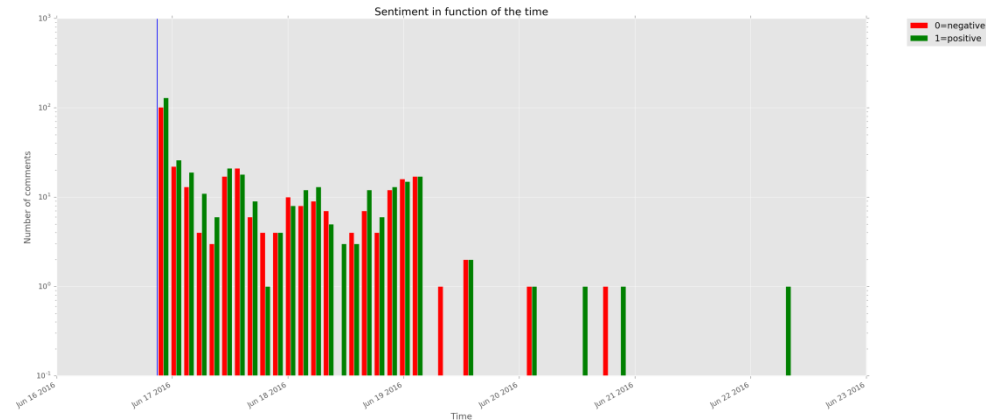
CNN articles about gun sales after Orlando shooting

Sentiment in function of the time

Number of comment expressing each sentiment
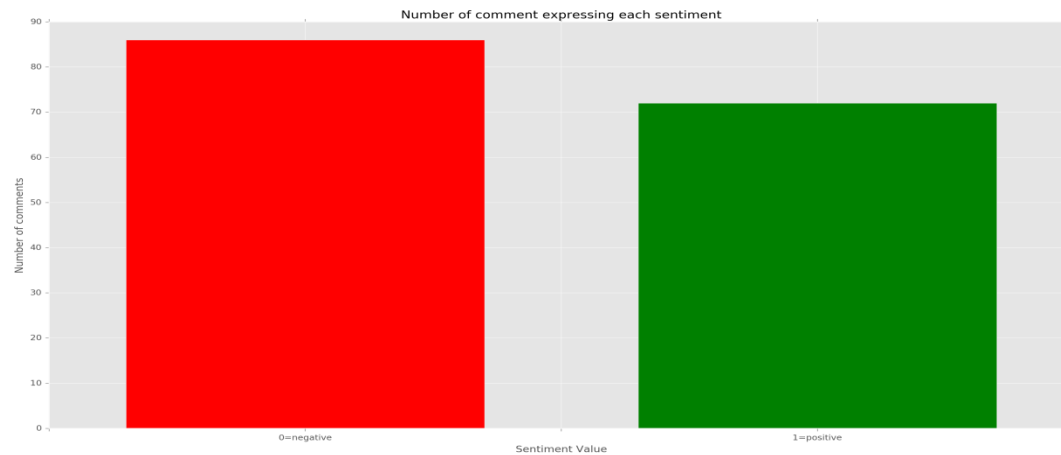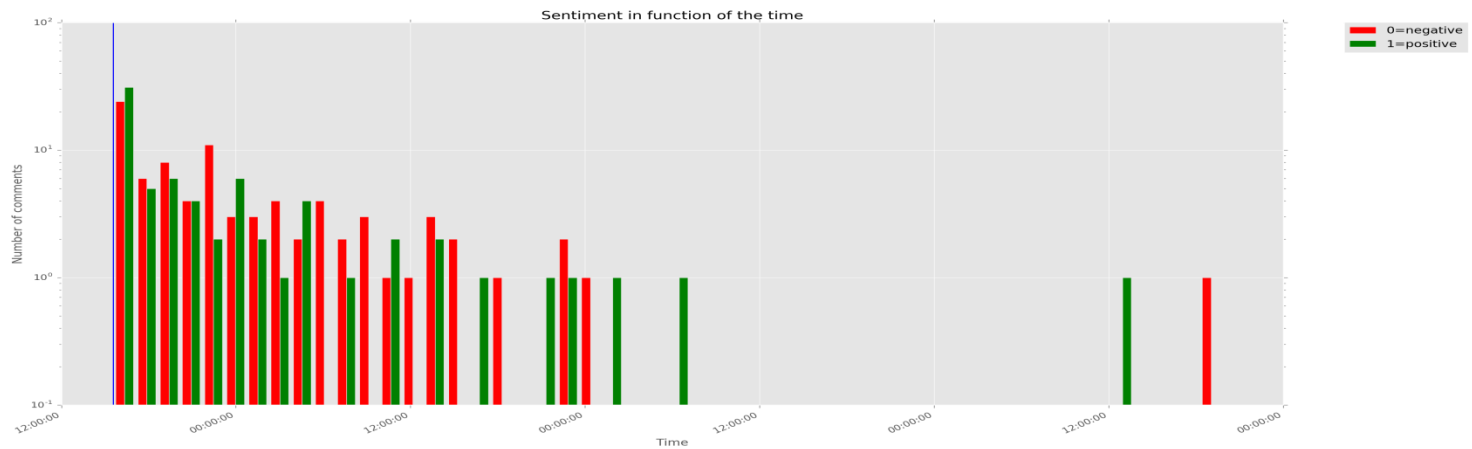
Theme: Abortion

CNN articles about Lena Dunham's abortion
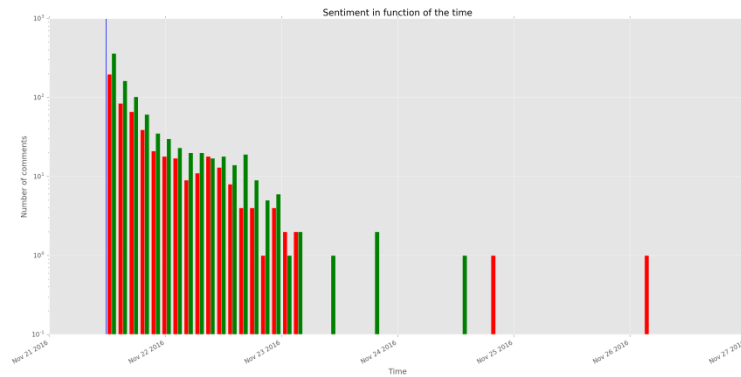
CNN article about Texas fetus burial obligation
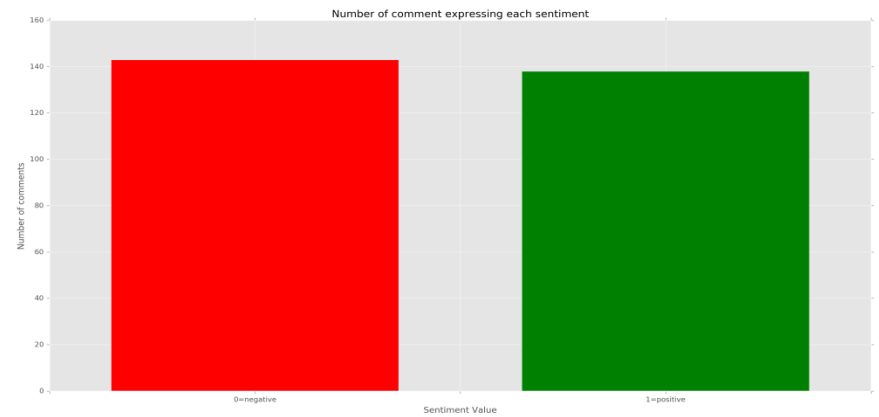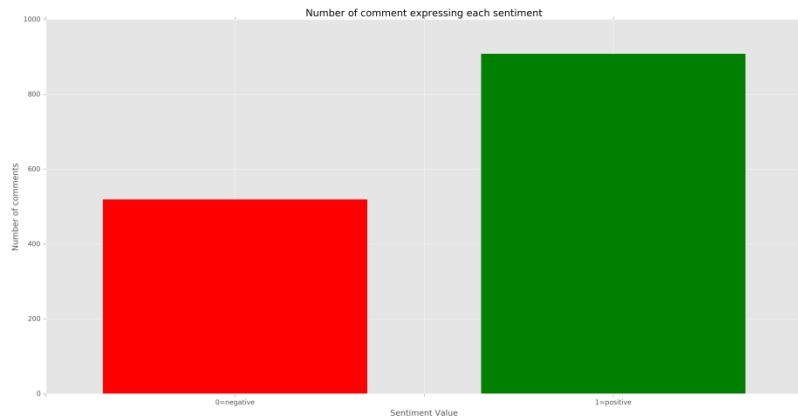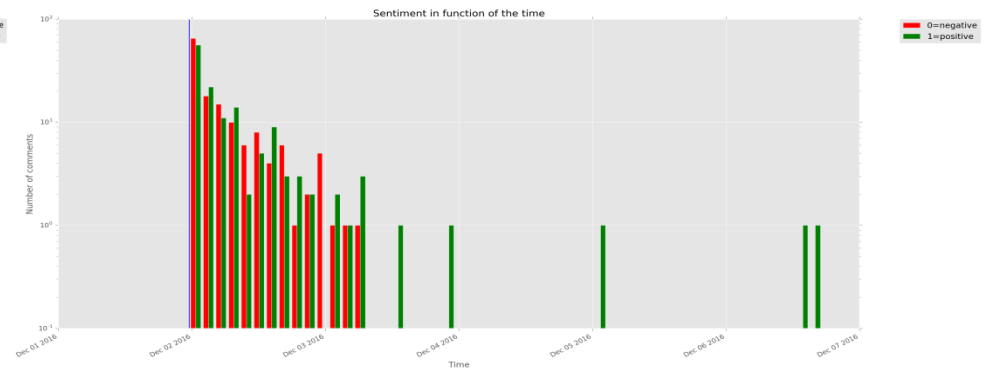
CNN article about pope abortion forgiveness
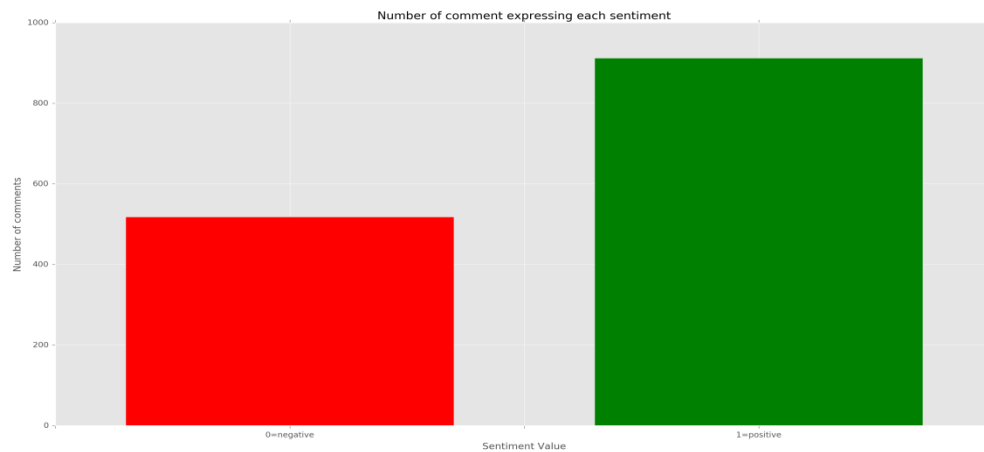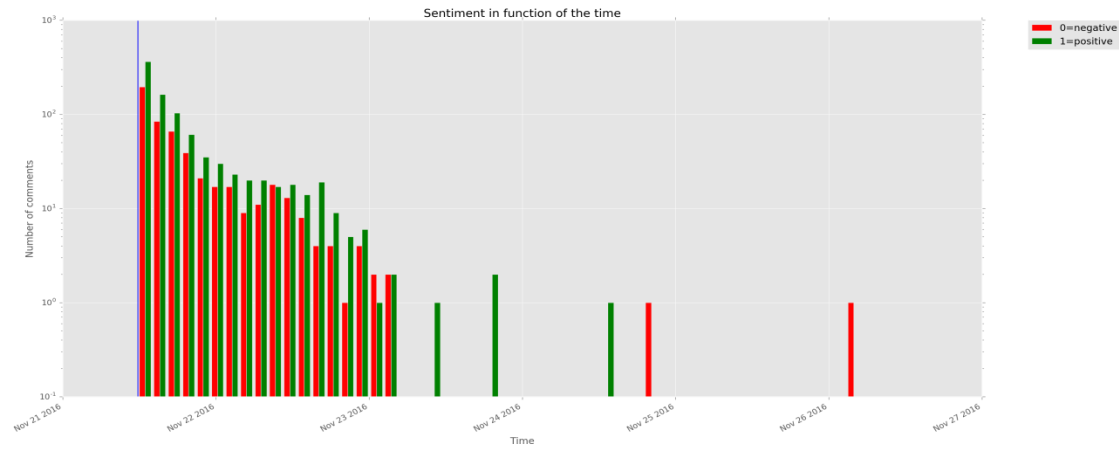
Theme: Political

David Cameron before brexit speaking about
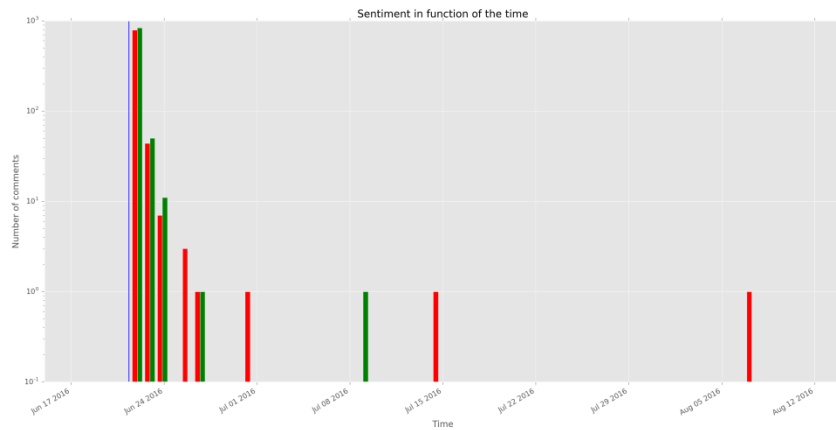
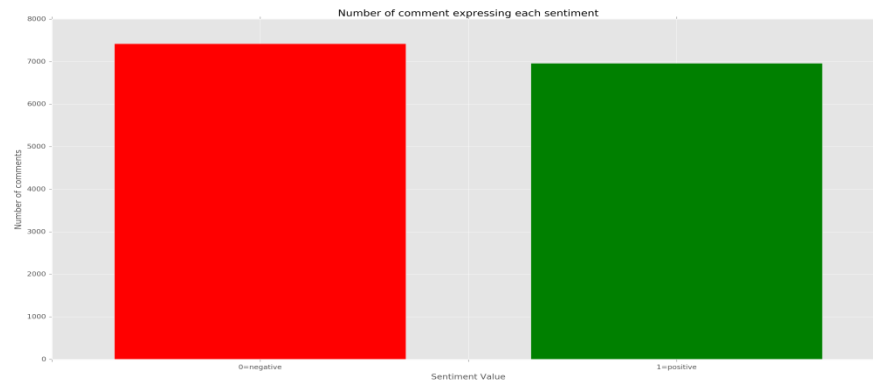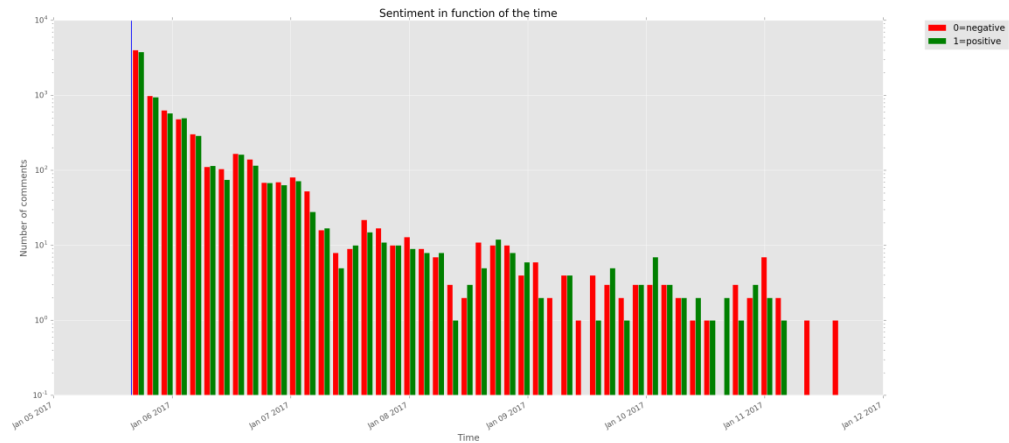**David Cameron 'thank you' note to the voters**
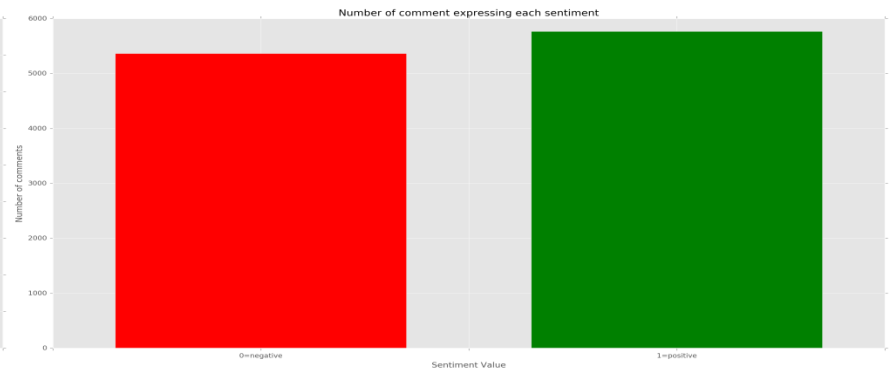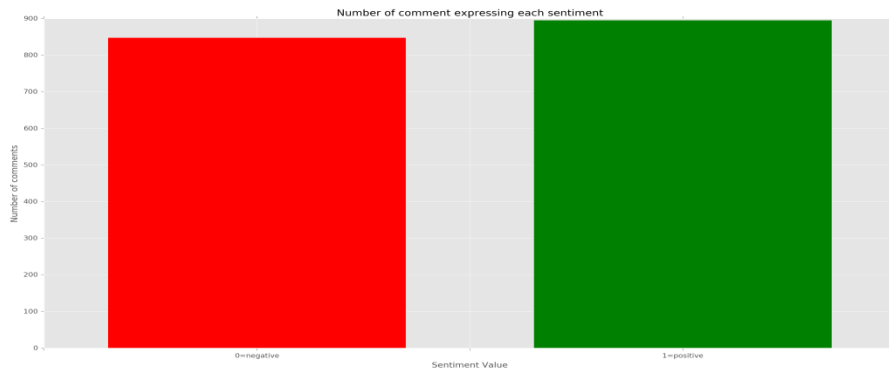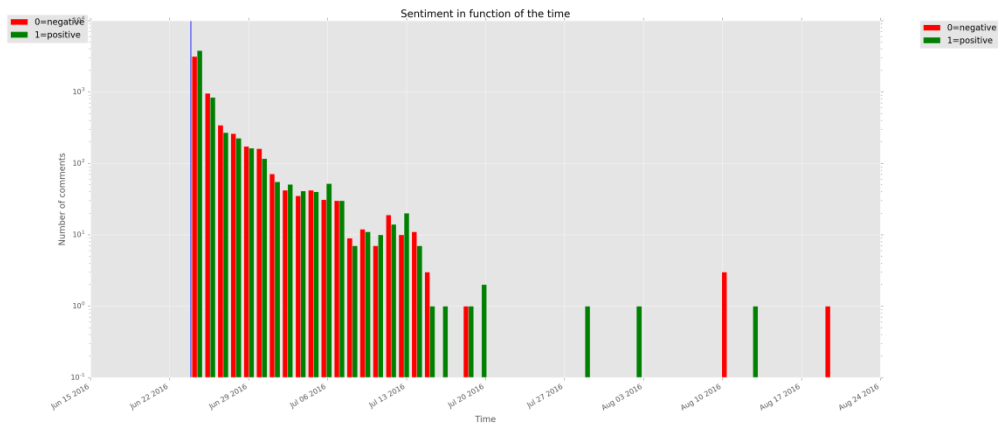
Theme: US Presidential Election

Hillary Clinton asking for votes (before election)
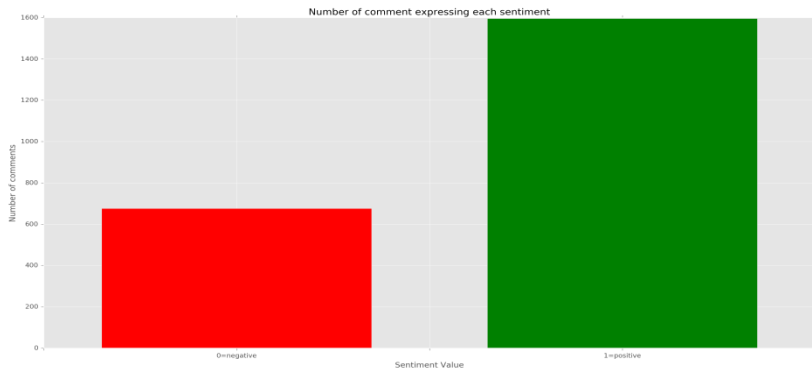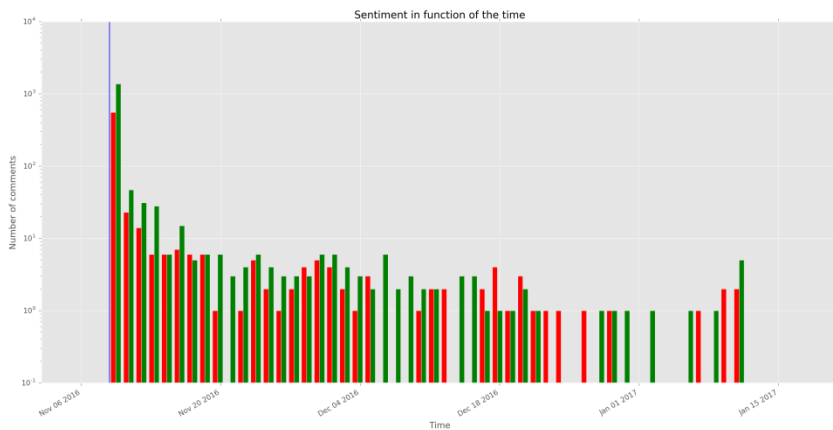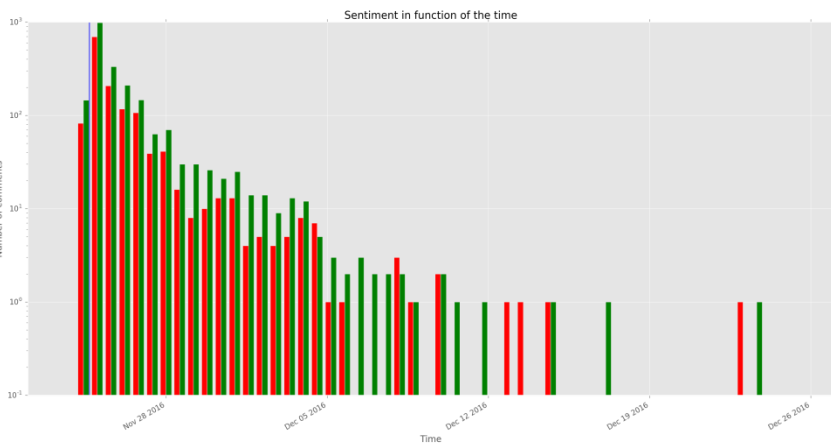
Donald Trump speaks about the defeat of the democrats
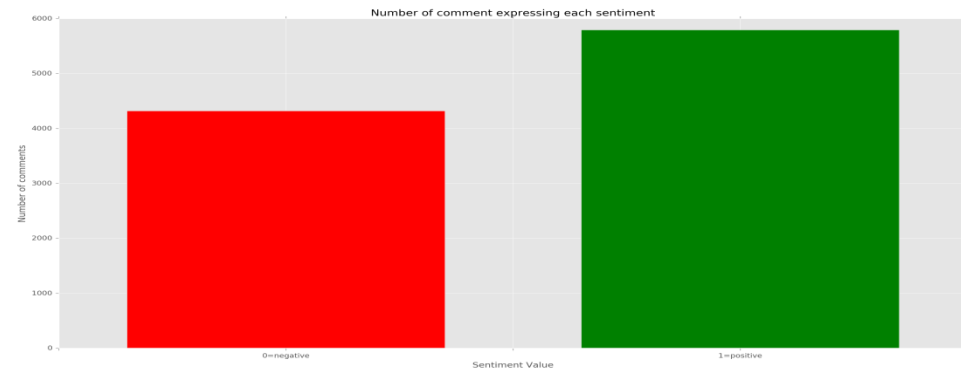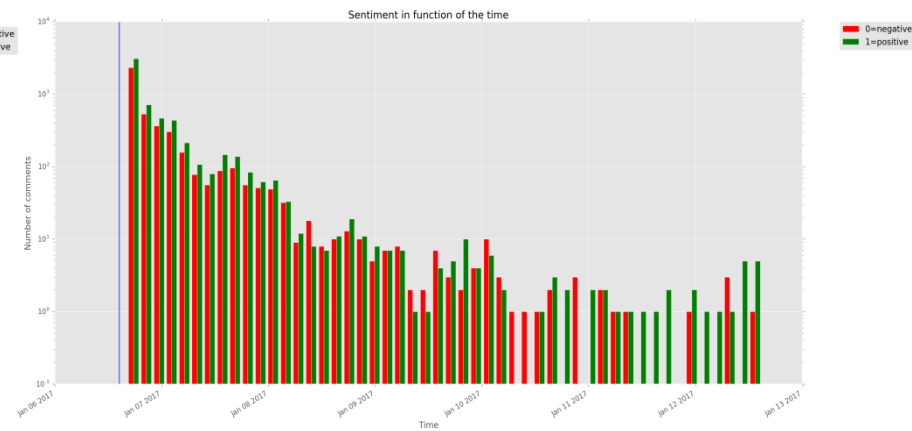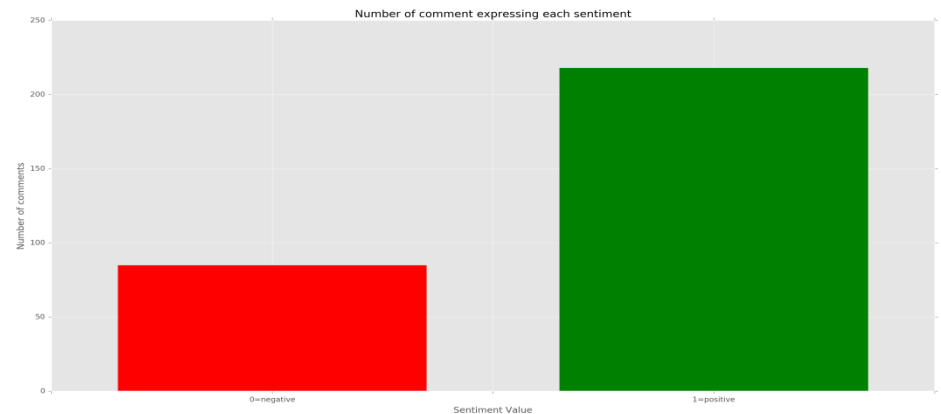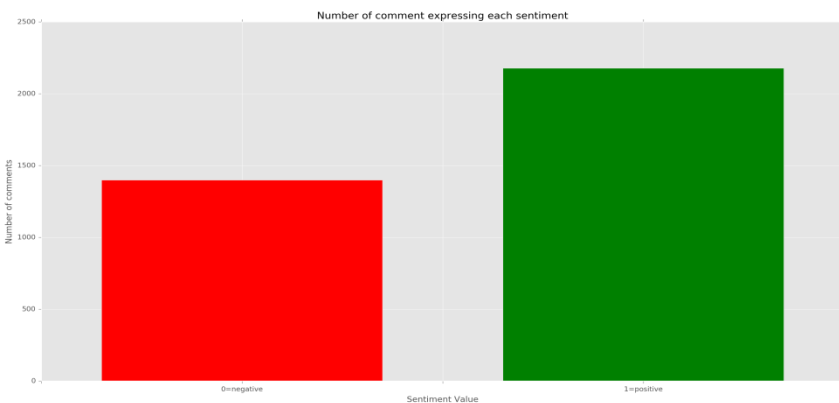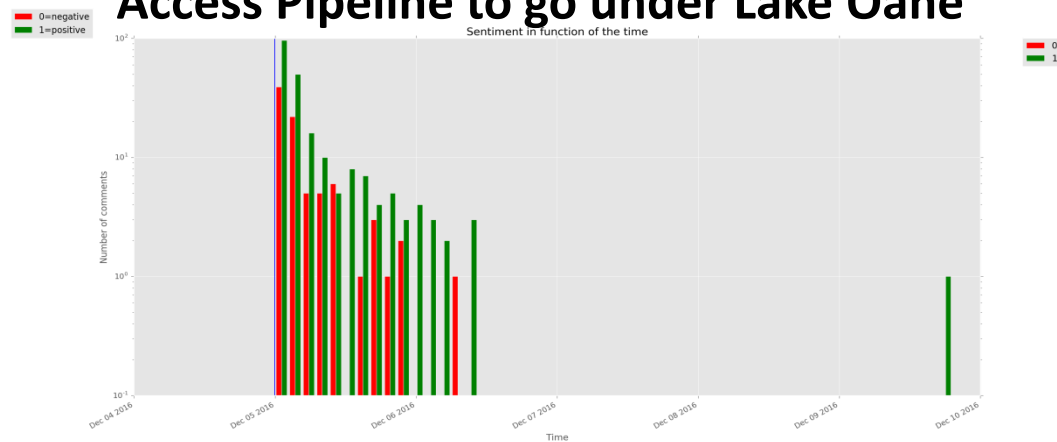
Donald Trump speaks about Obamacare

Theme: Environmental

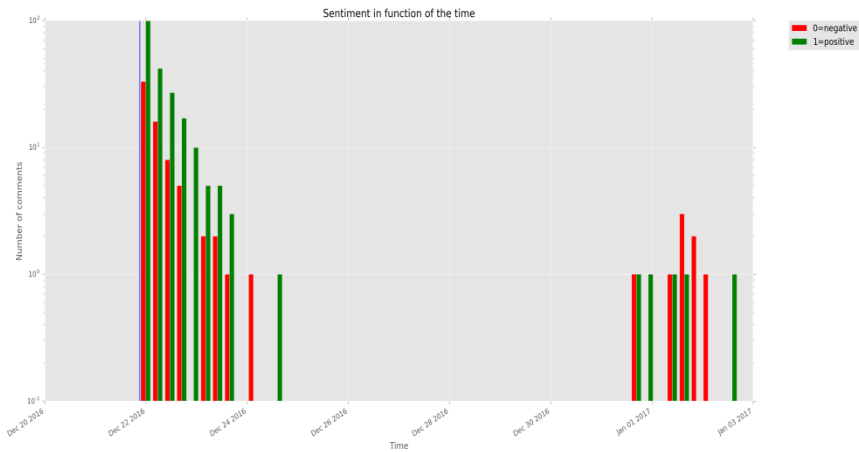**AJ+ post on Native Americans fighting against the Dakota Access pipeline**

AJ+ post on celebrations after the U.S. Army Corp of Engineers refused to grant an easement allowing the Dakota Access Pipeline to go under Lake Oahe
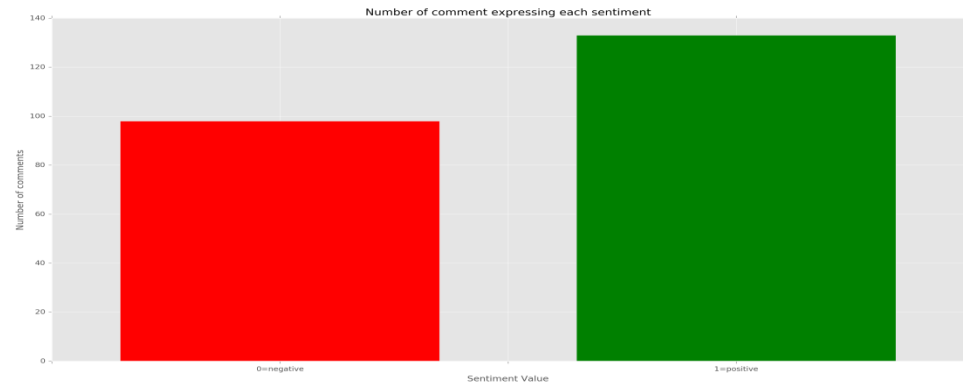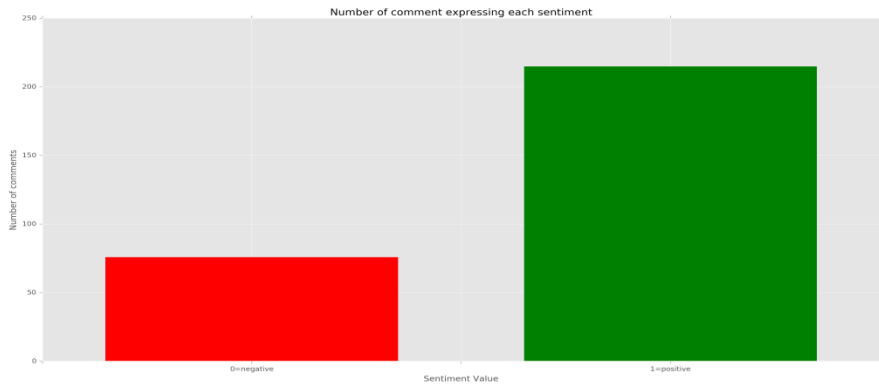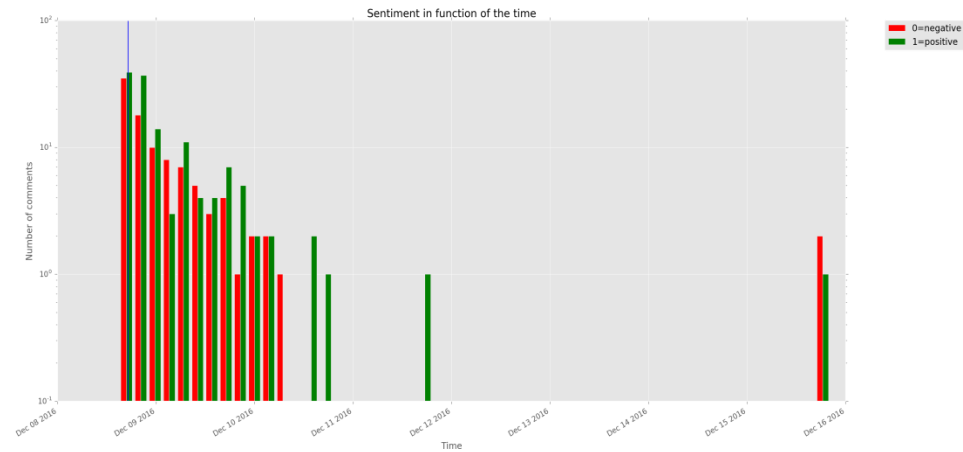
Donald Trump and Environmental Issues

AJ+ post on the hackers that are preserving environmental data before Trump takes office
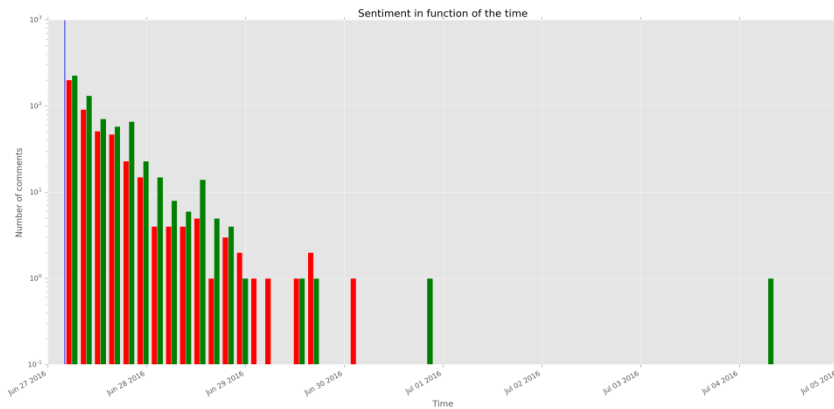
AJ+ post on Donald Trump being picked for the Environmental Protection Agency not believing in man-made climate change.
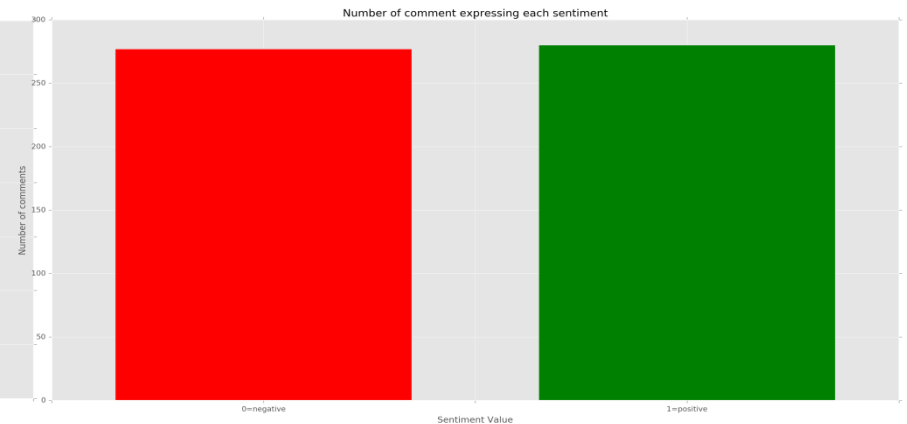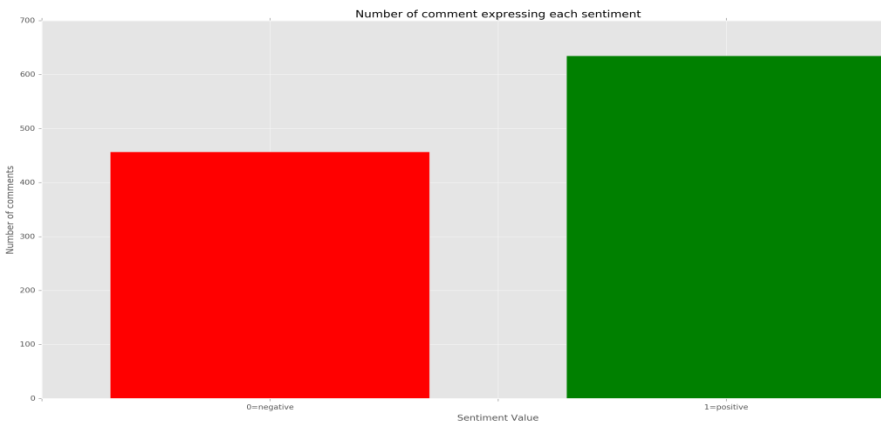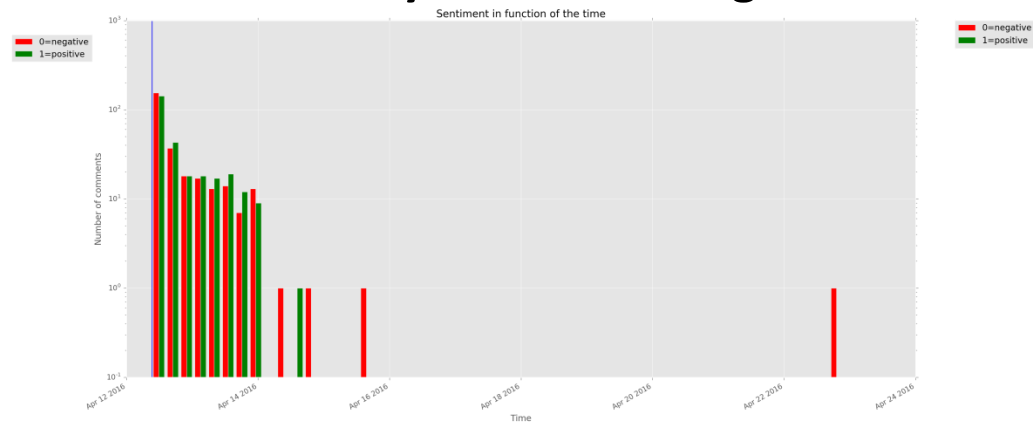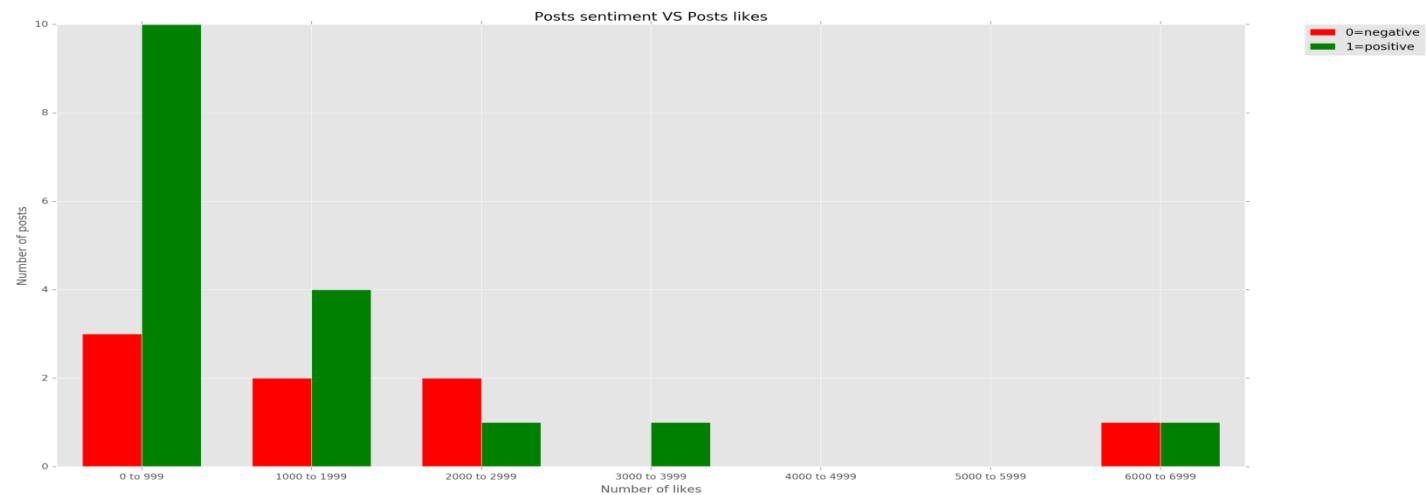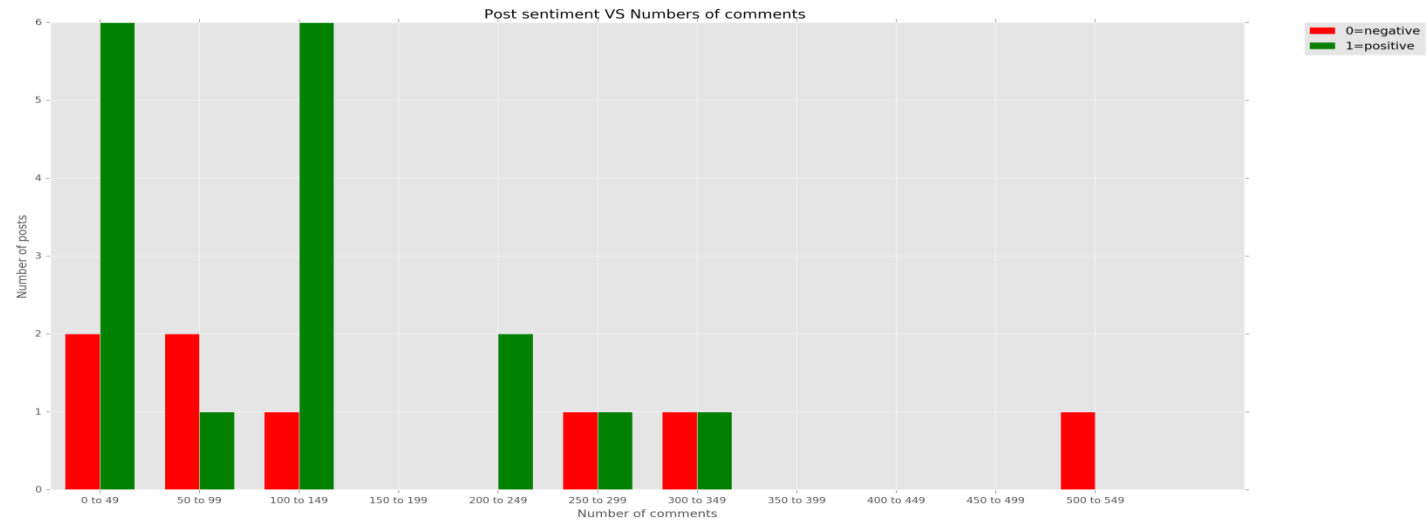
Theme: Homosexuality

CNN post on Pope saying Christians should apologize to gay people

CNN post on the fact that more than half of British Muslims think homosexuality should be illegal

Analysis on CNN Page

Questions?