Exploring and Visualizing Mental Health based on Features from Social Media

Johann Maximilian Zollner
maximilian.zollner@tum.de

Ramesse Zatti
ramesse.zatti@tum.de

Shou Shen
shou.shen@tum.de

Moritz Laass
moritz.laass@tum.de

Martin Werner
martin.werner@tum.de

Technische Universität München
Big Geospatial Data Management
Germany

Abstract

This paper presents a workflow for early detection of signs of depression from user-generated social media content. By analyzing text messages from Twitter and using supervised learning methods for classification, mental health issues can be detected to prevent serious health consequences. Syntactic feature, Bag of Words features and follower and friends count have been added to the training data-set and been considered for the classification. AB, LR, NB, RF and SVM are used for classification and the models achieve reasonable performance while the overall results evidence the importance of text features for depression detection. The results are visualized with ESRI ArcGIS to show the spatial distribution of the data.

1. Introduction

We present a system and workflow to detect signs of depression from user generated text messages which has been an active field in research recently. In this example we mine geolocated Twitter data from the area of New York as well as Washington and map features knowingly related to mental health based on the work fromRíssola et al.. [7] The data is classified with several different models and the results are visualized as maps. Finally, the results can be easily analyzed, compared and used for further studies.

2. Methodology

2.1. Data Retrieval and Processing

Twitter API v1.1 The standard API v1.1 was launched in 2012 and enables to retrieve data for resources such as Tweets, Users, Direct Messages, Lists, Trends, Media, and Places. For accessing the API with a Twitter developer account, a formal application to Twitter is required.

Data Mining Twitter data was mined with Python scripts running on a Jupyter Notebook. After a script connects to a stream, it continuously pulls data. The data streams are limited by bounding boxes, one for New York and one for Washington. Only geolocated Tweets within the bounding box are pulled but due to that only approximated 1% of the total data is geolocated, also Tweets without coordinates but instead a tagged Place like for example "NYC" were taken into account. The pulled data is saved in JSON format.

Preprocessing The JSON files containing the Tweets are preprocessed and transformed to CSV format. Preprocessing includes the following steps: extracting the relevant features like user, text and place; removing non-ASCII characters and replacing emojis; concatenating splitted files; removing pictures; extracting latitude and longitude; adding Bag of Words features as well as Syntactic features. All the preprocessing steps as well as the used scripts can be found in our LRZ Git repository.

2.2. Training Data

As training data, the labeled data-set "Depression Detection via Harvesting Social Media: A Multi-modal Dictionary Learning Solution" from Shen et al. [8] was used. The data-set "employed heuristical rule-based methods to construct two benchmark well-labeled depression and non-depression data-sets" [8] based on Twitter data. Six depression oriented feature groups were extracted and defined to describe each user: Social Network Feature, User Profile Feature, Visual Feature, Emotional Feature, Topic-level Feature and Domain-specific Feature. [8]

2.3. Features

Bag of Words Features In the bag of words model, a text is represented as the bag of its words, disregarding grammar and word order but keeping multiplicity. [2] After an analysis of the training data, the two most frequently used words were selected: from the depressed data the words "Depression" and "Diagnosed", while for the not-depressed data the words "Like" and "You". Additionally, the word "I" was selected, since it resulted strongly connected to depression as described by Rissola et al. [7] The frequency of these five words were counted for every Tweet and considered by the algorithm.

Syntatic Features Syntactic features [9] are formal properties of syntactic objects which determine how they behave with respect to syntactic constraints and operations (such as selection, licensing, agreement, and movement), which include AFINN score, syllable count, difficult words, Flesch Reading Ease (FRES) score, Flesch–Kincaid grade, Coleman–Liau index,automated readability index (ARI), Gunning fog index, Linsear Write formula, Dale–Chall readability formula, Crawford.

Social Engagement & Writing Behaviour Follower count, friend count, and favorite count are important indicators to measure users' social relationships. So we maintain them in our data.

2.4. Model Training

First, we randomly divide 60% of the data into training set, 20% evaluation set and 20% test set. Then, input the training sets in our five models:

- Adaptive Boosting(AB) [3]:is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances.
- Naive Bayes classifier (NB) [6]: assign the most likely class to a given example described by its feature vector.
- Logistic regression (LR) [4]:is used to model the probability of a certain class or event existing.
- Random Forest Classifier (RF) [5]: is an ensemble classification scheme that utilizes a majority vote to predict classes based on the partition of data from multiple decision trees.
- Support vector machine (SVM) [1]:is a linear model for classification and regression problems. bound constraints and one linear equality constraint.

2.5. Model Evaluation

We use the accuracy on the given data and labels to evaluate the models.

$$Accuracy = \frac{Number of correct prediction}{Total number of prediction}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives [10].

2.6. Prediction

After all model adjustments are completed, we can predict our unlabeled data and test data sets to provide an unbiased assessment of the final model fit on the training data set. In total 250155 tweets were binary classified for New York and 95271 for Washington.

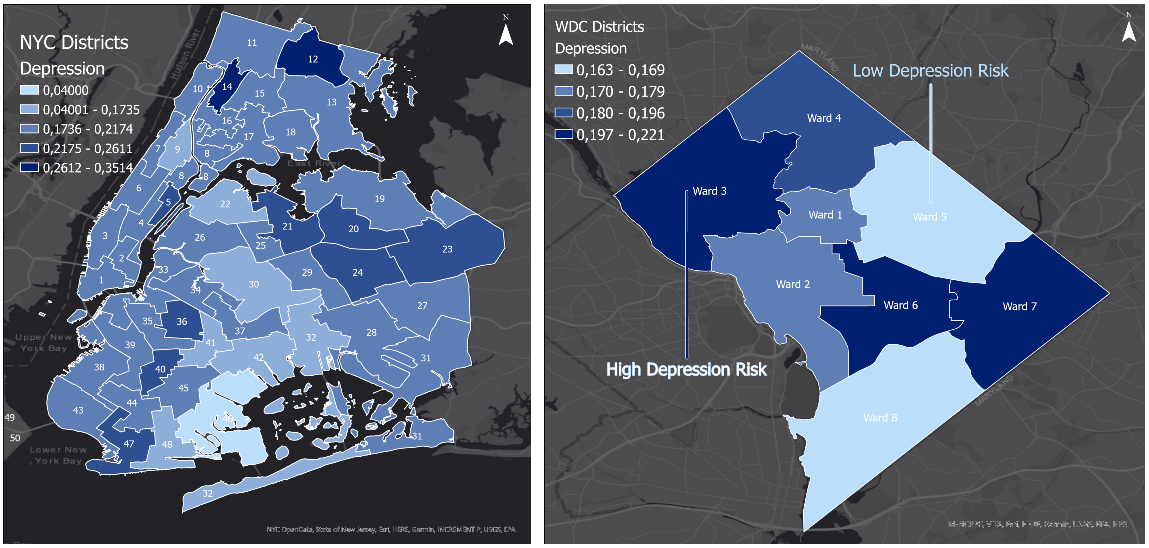
Number of Depressed tweets per Models		
Models	New York	Washington
0:	143264	56140
1:	31086	11630
2:	26078	9275
3:	25815	9040
4:	15232	6161
5:	8680	3025
Tweets:	250155	95271

Table 1. Number of Models that Classify a Tweet as Depressed

Table 1 shows the number of models that classified a tweet as depressed. For the later visualization of the data one tweet is considered depressed if at least 3 models classified it as depressed.

3. Results and Discussion

Table 2 shows the accuracy results of the five models. The accuracy of all models is above 0.7. The average accuracy rate is above 0.85. Linear regression and random forest models achieved the best results, with an accuracy of over 0.9.



(a) New York Area

(b) Washington Area

Figure 1. Normalized Depression Prediction Divided in Counties

Models			
Models	Train Accuracy	Evaluation Accuracy	Test Accuracy
AB	0.849	0.840	0.847
LR	0.911	0.894	0.904
NB	0.736	0.730	0.744
RF	1.00	0.925	0.925
SVM	0.887	0.887	0.880
average	0.876	0.855	0.860

Table 2. Models' Accuracy

A collection of maps was designed to show the spatial distribution of risk of depression in New York and Washington, based on geolocated tweets. For this visualization, only tweets with coordinates were used. The New York maps represent 14705 geolocated tweets, while the Washington maps only represent 3163 tweets due to less twitter activity in this area.

Figure 1 represents the prediction for risk of depression from the twitter data divided in the different council districts. This representation can, for example, be useful to find out which parts of the represented cities need more psychological support.

Figure 2 is generated with the Kriging interpolation method and show hot-spots of depression around the cities. The New York map result is smoother and more representative, since the number of data points is considerably higher.

Figure 3 presents the relation between the depression prediction (represented by colors) and the the total follower count (represented by the height) for every council district. This representation does not reveal any particular insights, but demonstrates the possibility of further visualizations.

4. Conclusions

We presented a system and workflow to detect signs of depression from user generated social media content. Multiple features were extracted to train models based on the training data from Shen et al. [8]. Different models were evaluated and an average accuracy of 0.85 was achieved. A collection of maps was designed to show the variety of possibilities of the representation of the data. The results can be easily analyzed, compared and used for further studies.

5. Remarks

All used data, scripts, results and the basic proposed workflows can be found in our Git repository "bgd_mentalhealth" on gitlab.lrz.de .

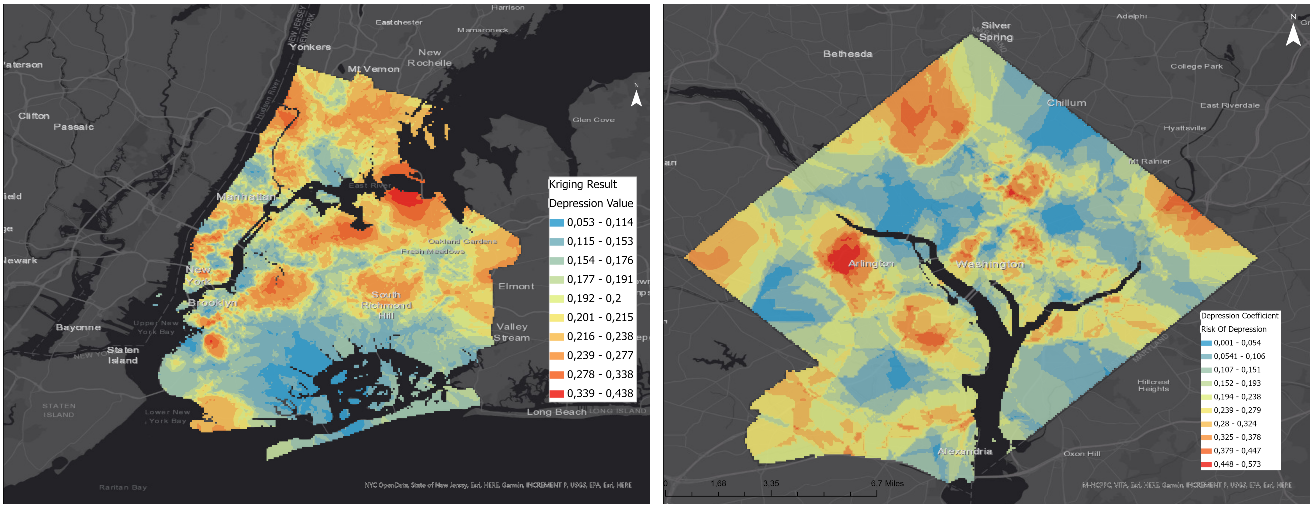
References

[1] T. Joachims. Making large-scale svm learning practical. Technical report, Technical report, 1998.

[2] S. Josef. Efficient visual search of videos cast as text retrieval. *IEEE TRANSACTIONSON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 31:591–605, 2009.

[3] D. D. Margineantu and T. G. Dietterich. Pruning adaptive boosting. In *ICML*, volume 97, pages 211–218. Citeseer, 1997.

[4] S. Menard. *Applied logistic regression analysis*, volume 106. Sage, 2002.



(a) New York Area (b) Washington Area

Figure 2. Kriging Interpolation of Depression Prediction

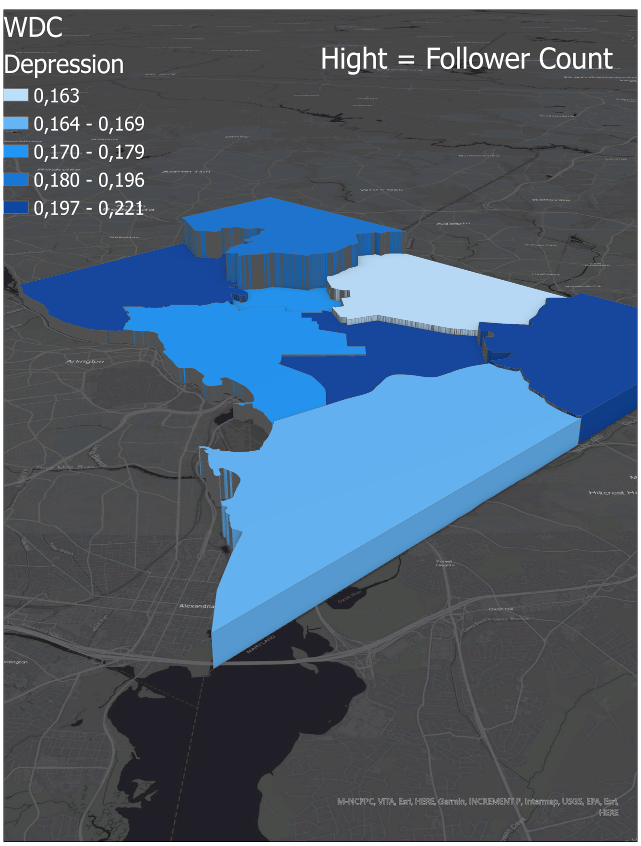


Figure 3. Relation Between Depression Prediction and Follower Count Divided in Counties of the Washington Area

[5] M. Pal. Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1):217–222, 2005.

[6] I. Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[7] A.Ríssola, D. E. Losada, and F. Crestani. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare*, 2(17):1–31, March 2021.

[8] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844, 2017.

[9] P. Svenonius. Syntactic features, 04 2019.

[10] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap. Confusion matrix-based feature selection. *MAICS*, 710:120–127, 2011.