

OBJECT-ORIENTED APPROACH FOR ETHNIC ENMITY DETECTION IN TEXT MESSAGES BY NLP

Molchanova Maryna

Teacher

m.o.molchanova@gmail.com

Mazurets Oleksandr

Ph.D (Engineering Science), Associate Professor

exe.chong@gmail.com

Sobko Olena

Teacher

olena.sobko.ua@gmail.com

Boiarchuk Illia

Postgraduate student

romaboy2005@gmail.com

Computer Science Department

Khmelnyskyi National University, Ukraine

In recent years, social media has witnessed a surge in popularity among users worldwide, transforming into a crucial channel for expressing opinions, views, and emotions [1]. However, conflicts and tensions frequently arise between different ethnic groups globally. These situations often manifest in expressions on social media, necessitating the identification and analysis of such expressions [2].

The identification of ethnic hatred manifestations is a crucial societal issue, as it can lead to severe consequences, including conflicts and divisions between different ethnic groups [3].

Natural language processing (NLP) tools are gaining significant traction in the field of text analysis [4]. They enable the automation and facilitation of large-scale data analysis, encompassing texts published on social media platforms [5].

The aim of this work is to develop a software implementation of a method for detecting manifestations of ethnic hatred in text messages of social Internet networks using NLP tools.

The method for identifying ethnic hatred manifestations in social media text messages using NLP tools is designed for the automated analysis of texts published on social media platforms to detect signs of hatred or conflict between representatives of different ethnic groups. The method employs natural language processing techniques, namely an ensemble-based approach, and transforms the input data in the form of a trained FastForest classifier and an input text message into the output data in the form of the percentage of ethnic hatred manifestation in the test social media message. The method's diagram and steps are presented in Figure 1.

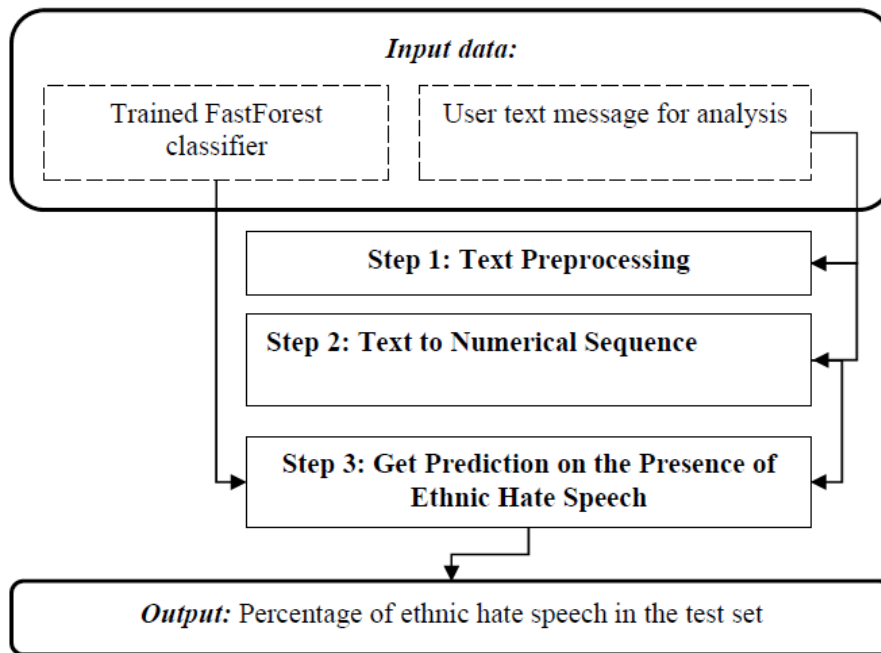


Figure 1. Schematic Diagram and Steps for Detecting Ethnic Hate Speech in Text Messages.

The inputs to the method for detecting manifestations of ethnic hostility in text messages are a trained FastForest classifier and a custom text message for analysis.

The work of the method is divided into 2 main stages - pre-processing of the user message and prediction by the FastForest classifier. The first step is the preprocessing step, which includes removing stop characters, removing stop words, and removing emoticons.

The second step is to convert the text into a numeric sequence, which will be the input to the trained FastForest machine learning model.

The next step is to make a prediction about the presence of ethnic hatred based on the FastForest machine learning model, which is loaded with the numerical sequence from the previous step.

The output of the method is the percentage of ethnic hatred in the test message.

The .NET 7.0 platform, the Visual Studio programming environment, and the C# programming language will be used to write a software implementation of the application to study the effectiveness of the method for detecting manifestations of ethnic hatred in text messages on social Internet networks.

To ensure the functionality of the information system for detecting manifestations of ethnic hatred, which consists of 3 main subsystems: "Subsystems for working with experimental data", "Subsystems for detecting manifestations of ethnic hatred", "Subsystems for preliminary data processing"; 2 auxiliary subsystems: "Subsystem for generating a training sample from the dataset" and "Subsystems for training a machine learning model", the corresponding class diagram was designed (Figure 2).

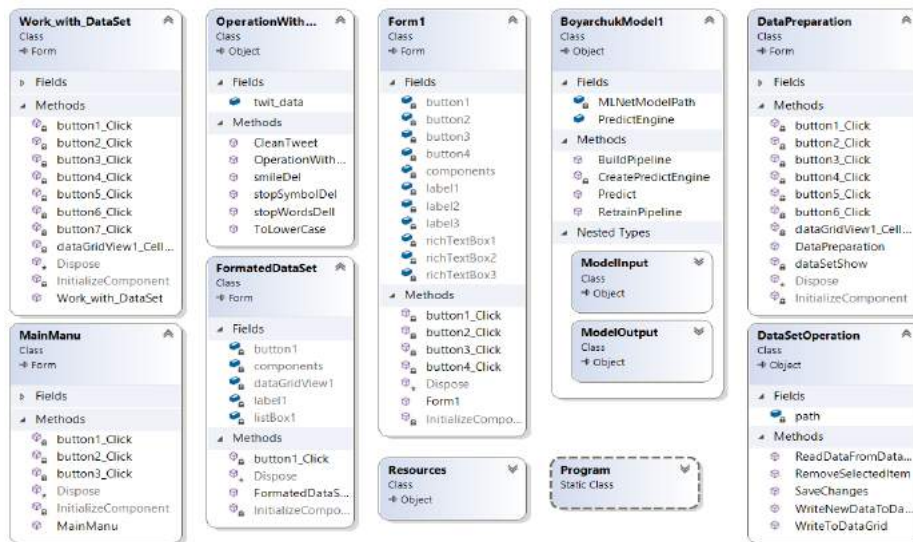


Figure 2. Application class diagram.

The machine learning model training subsystems are implemented by the corresponding class "BoyarchukModel1", which has methods for building a pipeline, training a neural network, and saving the result. The result of the subsystem is a trained FastForest machine learning model.

The "FormatedDataSet" class is intended to implement an auxiliary subsystem for forming a training sample from the dataset and displaying statistics on the available categories in the dataset. The result of the subsystem is a formed working dataset that will be used as input for other subsystems.

The subsystems for working with experimental data are implemented using the "Work_with_DataSet" and "DataSetOperation" classes, which contains the ReadDataFromDataSet() methods for reading data from the dataset WriteNewDataToDataSet() to write new data to the dataset, the SaveChanges() method is designed to save changes after working with the working data, and RemoveSelectedItem() to remove the selected record from the working data.

The data preprocessing subsystems are implemented by the functionality of the "DataPreparation" and "OperationWithData" classes. The dataSetShow() method is designed to display the working dataset in a list, in the button click handlers, methods of the "OperationWithData" class are called. The smileDel() method is designed to remove emoticons. The stopWordsDell() method is designed to remove stop words. The stopSymbolDel() method is designed to remove stop characters. The CleanTweet() method is designed to clean the text from all types of information noise at once.

The subsystems for detecting manifestations of ethnic hatred are implemented by the "Form1" class. The main functionality of the subsystem is implemented in the event handlers of button clicks. The input data is the trained FastForest model, and is the main subsystem of the designed application.

The lengths of tweets are determined, each tweet is analyzed, and for each record its length in characters is calculated. The data is saved to a text file with the .txt extension. The view of statistics on records of date sets is shown in Figure 3.

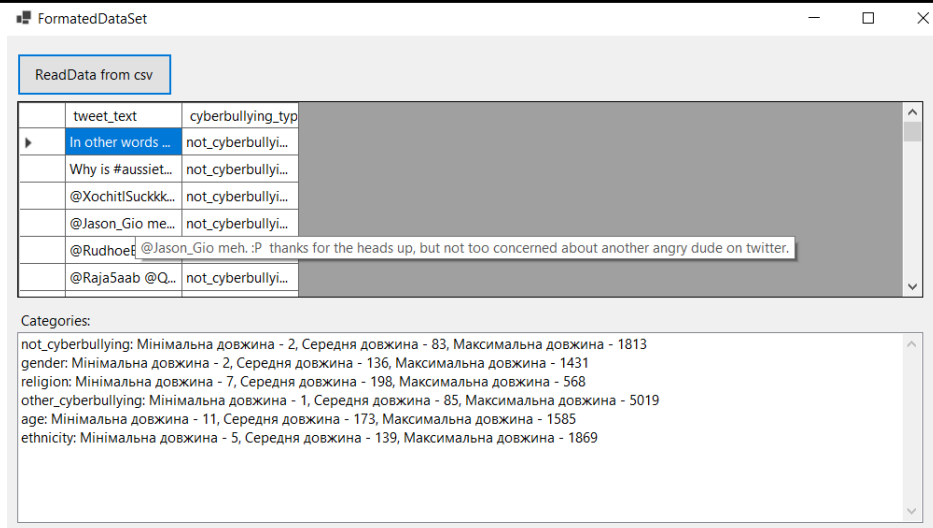


Figure 3. Subsystem for generating a working dataset.

The main subsystem is the ethnic hatred detection subsystem, which uses a trained FastForest machine learning model. To train the model, the BuildPipeline() method takes an MLContext object as a parameter and returns a pipeline for data processing. The pipeline defines a sequence of data transformations that will be applied to the input data before training the model. The trained model is saved for further use.

To determine whether there are manifestations of ethnic hatred, an instance of sampleData of the class "BoyarchukModel1.ModelInput" is created, which represents the input data for the model. The value for analysis is taken from the text contained in the text field. The Predict method from the BoyarchukModel1 class is used to make a prediction based on the sampleData input. The result of the prediction is saved in the predictionResult variable. The implementation of the described functionality is shown in Figure 4.

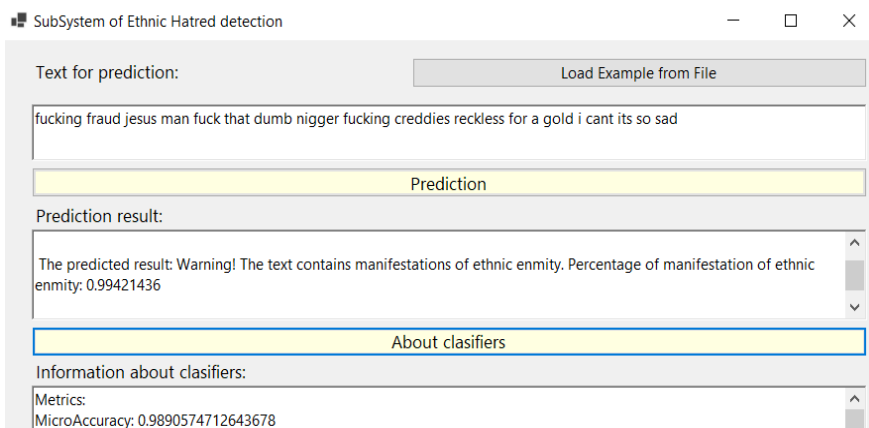


Figure 4. An example of identifying manifestations of ethnic hatred

The effectiveness of the method was studied using the developed software by comparing the obtained answers with the validation set, and the trained FastForest machine learning model was evaluated using the metrics MicroAccuracy, MacroAccuracy, LogLoss, ConfusionMatrix, f1-measure and Recall.

Without changing the working training set, the metrics values were as follows: MicroAccuracy 0.9890, MacroAccuracy 0.9889, LogLoss 0.0463.

Therefore, we have developed a software implementation of the method for detecting manifestations of ethnic hatred in text messages of social Internet networks by NLP tools, which uses natural language processing techniques and converts input data in the form of a trained FastForest classifier and an input text message into output data in the form of a percentage of ethnic hatred in a test message of social Internet networks.

References

1. Familylives. Bullying on social networks. URL: <https://www.familylives.org.uk/advice/bullying/cyberbullying/what-to-do-if-you-re-being-bullied-on-a-social-network>
2. Social Media Management. User roles. URL: <https://social-media-management-help.brandwatch.com/hc/en-us/articles/4626232528029-User-Roles>
3. Slobodzian V., Molchanova M., Kovalchuk O., Sobko O., Mazurets O., Barmak O., Krak I. An Approach Based on the Visualization Model for the Ukrainian Web Content Classification. 2022 12th International Conference on Advanced Computer Information Technologies, ACIT 2022. 2022. pp. 400-405.
4. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. CEUR Workshop Proceedings, 2023, vol. 3387, pp. 344-356.
5. Nazarov V., Molchanova M. Information System for Detecting Abusive Speech in Audio Content by Means of Natural Language. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. Pp. 132-135.

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ТЕСТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Зеленков Андрій

к.т.н., доцент

Кафедра Інженерії програмного забезпечення

a.zelenkov@khai.edu

Національний аерокосмічний університет ім. М.С. Жуковського

«Харківський авіаційний інститут», Україна

У сучасному світі більшість підприємств та мільйони приватних осіб щоденно використовують різне програмне забезпечення (ПЗ) для задоволення бізнесових та особистих потреб. Програмне забезпечення має бути якісним та надійним. Прояви дефектів у ПЗ можуть дратувати користувачів та змушувати