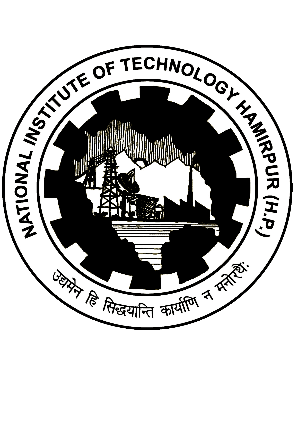
**NATIONAL INSTITUTE OF TECHNOLOGY HAMIRPUR**

**HAMIRPUR (H.P) – 177 005 (INDIA)**

****

**INTERNSHIP REPORT**

**ON**

**Fake News Detection Using LSTM**

(B.Tech 4th Year CSE NITH)

Submitted By:

* **Gajendra Surariya (185082)**
* **Kashish Srivastava (185014)**

Submitted To:

**Dr. Mohit Kumar**

July 20, 2021

**Department of Computer Science and Engineering,**

**NIT Hamirpur**

# INDEX

|  |  |  |  |
| --- | --- | --- | --- |
| S.No. | Contents | Page No. | |
| 1 | Acknowledgement | | 1 |
| 2 | Abstract | | 2 |
| 3 | Introduction | | 3 |
| 4 | Methodology | | 8 |
| 5 | Conclusion | | 13 |

# ACKNOWLEDGEMENT

On this great occasion of accomplishment of our project on **Fake News Detection**

**Using LSTM**, we would like to sincerely express our gratitude to **Dr. Mohit Kumar**, who has been supported through the completion of this project.

|  |
| --- |
| We would also be thankful to our Director **Dr.** **Lalit Kumar Awasthi** |

of **National Institute of Technology, Hamirpur** for providing all the required facilities in completion of this project.

Finally ,we would like to appreciate all our group members for their support and coordination, we hope we will achieve more in our future endeavors.

Kashish Srivastava,

Gajendra Surariya

# ABSTRACT

Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enables the wide spread of “Fake News”, i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Because the issue of fake news detection on social media is both challenging and relevant, we conducted this survey to further facilitate research on the problem. In this survey, we present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. We also discuss related research areas, open problems, and future research directions for fake news detection on social media.

# INTRODUCTION

During our Internship at National Institute of Technology, Hamirpur under Dr. Mohit Kumar, we were in a team of two interns. Together we worked on the problem stated above i.e. Detecting Fake News and labelling them and providing a place displaying deception free news and promoting trusted news. In order to achieve this goal we made a machine learning Model for Sequence classification using LSTM using Keras API in Tensorflow that classify news as fake or not fake i.e. trusted or not trusted. The model clearly identifies fake news. Also, we made social online platform to aggregate news from local reporters and for displaying bluff free News using Django framework. We implemented Authentication system to authenticate the authors identity and implemented various functionalities. We provided authors with an interactive user interface for adding news article and the author is given permission to delete the news article posted by them on the platform. There is also user interface where author can see all the news articles posted by them on the platform.

# METHODOLOGY

Our project has been divided into 2 subcomponents:

1. Sequence Classifier LSTM Model
2. Django Website

## Sequence Classifier LSTM Model

Sequence Classifier development has two major steps:

* Data pre-processing and cleaning
* Classifier Model Structure and Training

### Data Pre-processing and Cleaning

* Filling NULL values with empty string

All the NULL values in the dataset must be replaced with empty string as NULL values cannot be processed and used in training model.



* Removed URLs, emojis and punctuations

URLs, emojis and punctuations must be removed before processing text as they may temper with the results

* Tokenized base text and title

Tokenization is a common task in Natural Language Processing. Tokenization is a way of separating a piece of text into smaller units called tokens. Here, tokens can be either words, characters, or subwords.

* Lower cased clean text

Text is converted to lower case so that case does not affect the prediction.

* Removed stopwords

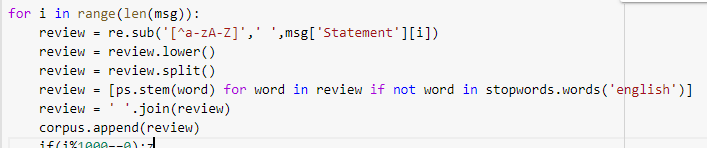
Stopwords are the English words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the

sentence.

* Applying word Stemming

**Stemming** is the process of reducing a **word** to its **word** stem that affixes to suffixes and prefixes or to the roots of **words** known as a lemma. **Stemming** is important in natural language understanding

All of the above transformations are applied to each string in dataset and saved in corpus.

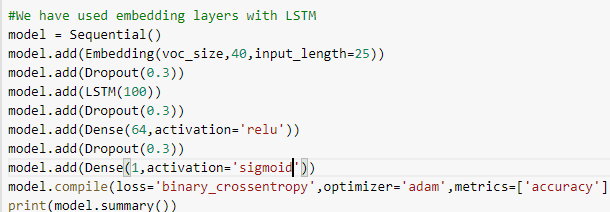


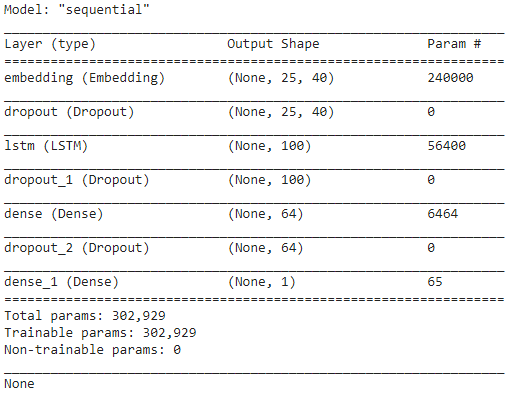
### Classifier Model Structure and Training

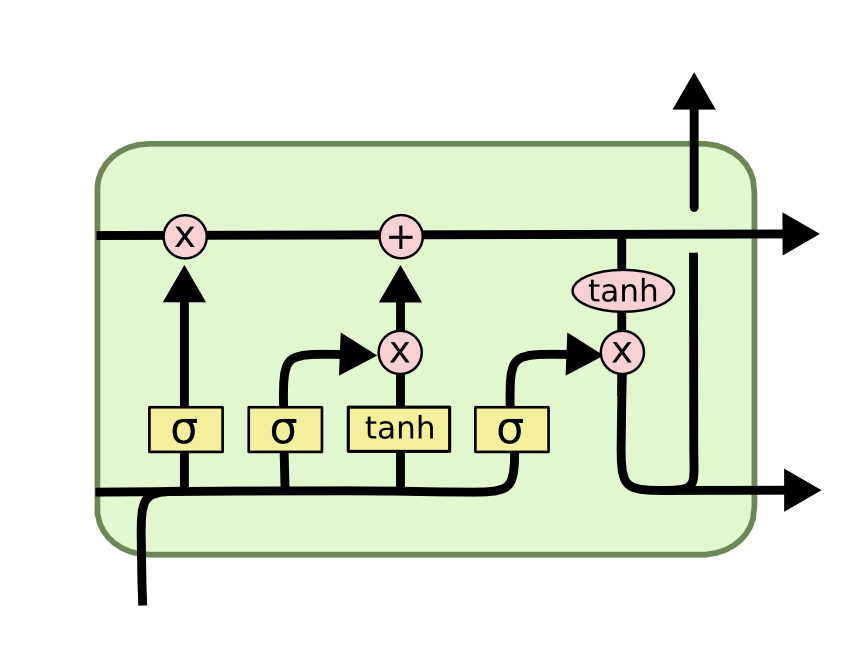
After pre-processing, cleaning and transforming the dataset to the inputs that can be used by our model we develop classifier model using keras over Tensorflow using the embedding layer with LSTM.

After developing the model we need to compile our model with following parameters:

* Loss Function: Binary Cross-Entropy
* Optimizer Function: Adam Optimizer
* Metrics: Accuracy







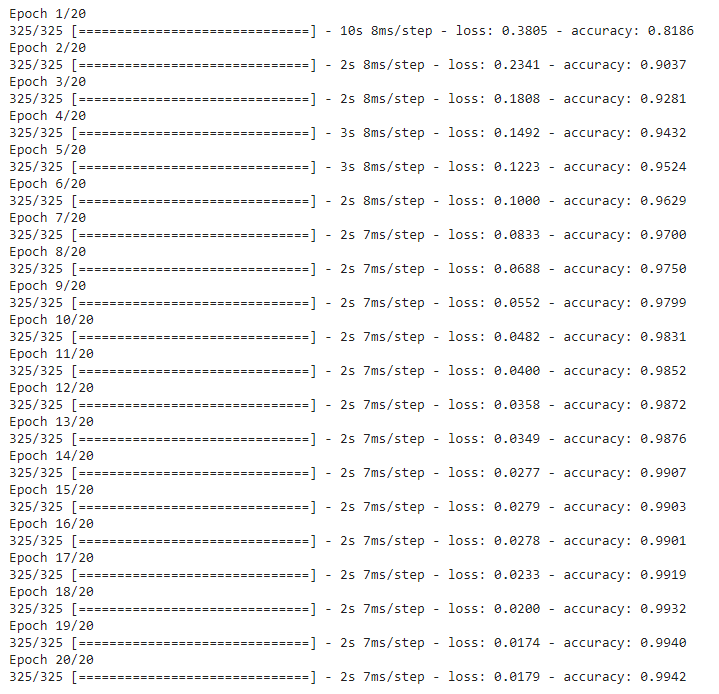
LSTM Structure

Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points, but also entire sequences of data, that’s why we prefer LSTM for text/sequence classification.

Now, we train our classifier using the embeddings with following parameters:

* Epochs: 20
* Batch size: 64

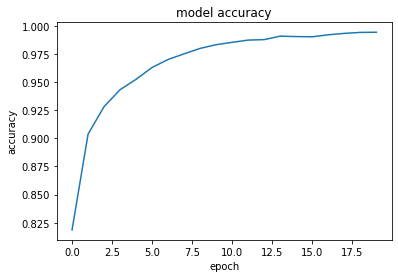


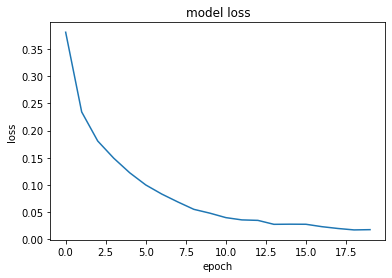


Model Training for 20 epochs

After training the model for 20 epochs we observe the testing accuracy of **99.42%** and a loss of **0.0179**.

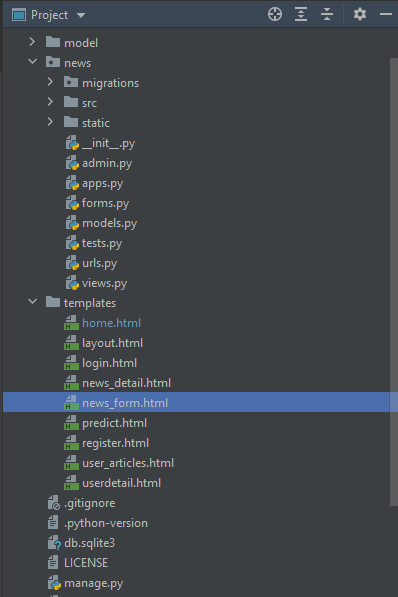
Following figures represents the model accuracy/loss relationship with each epoch:





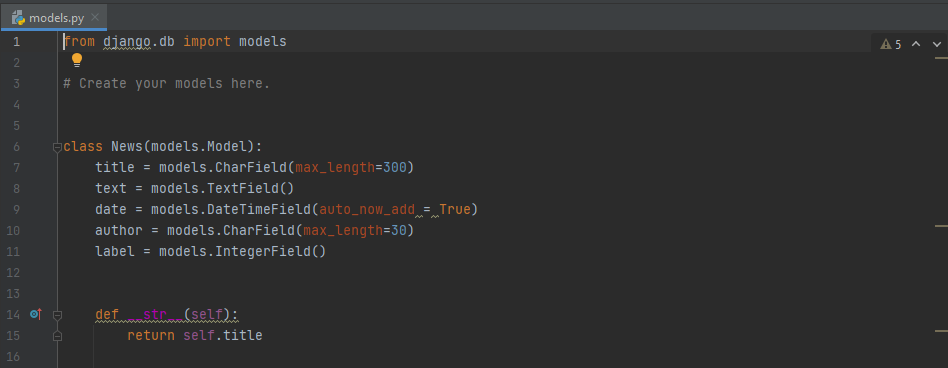
## Django Website

Our Django project has following directory structure.

.****

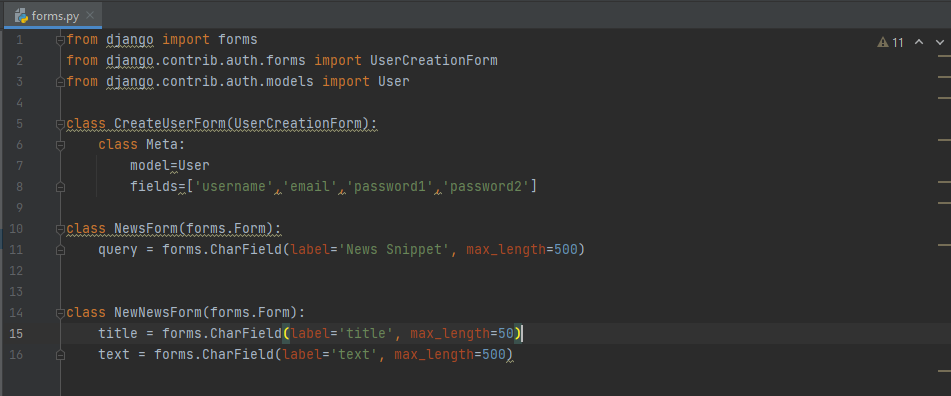
Here the models.py contains the models used in the project to store data.

Below is the code snippet of model.py

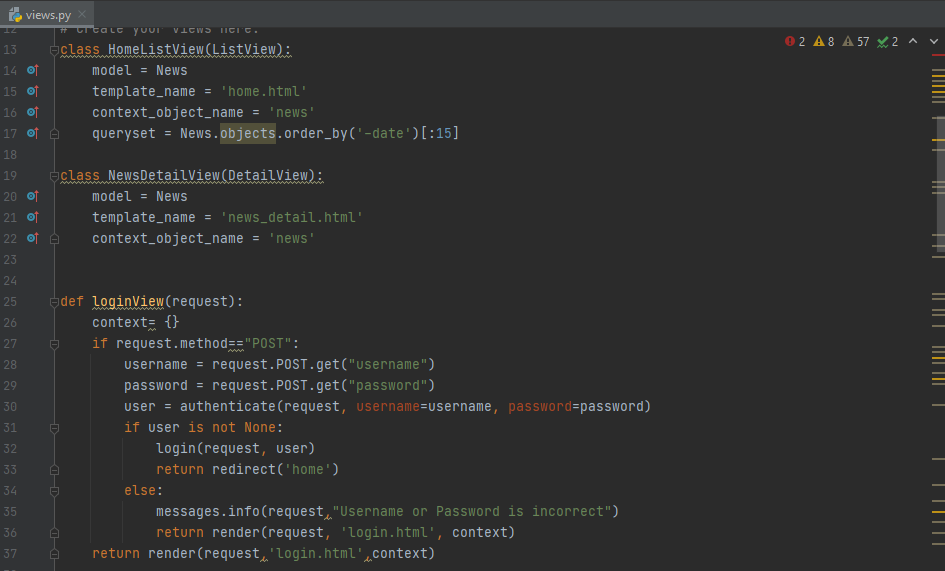


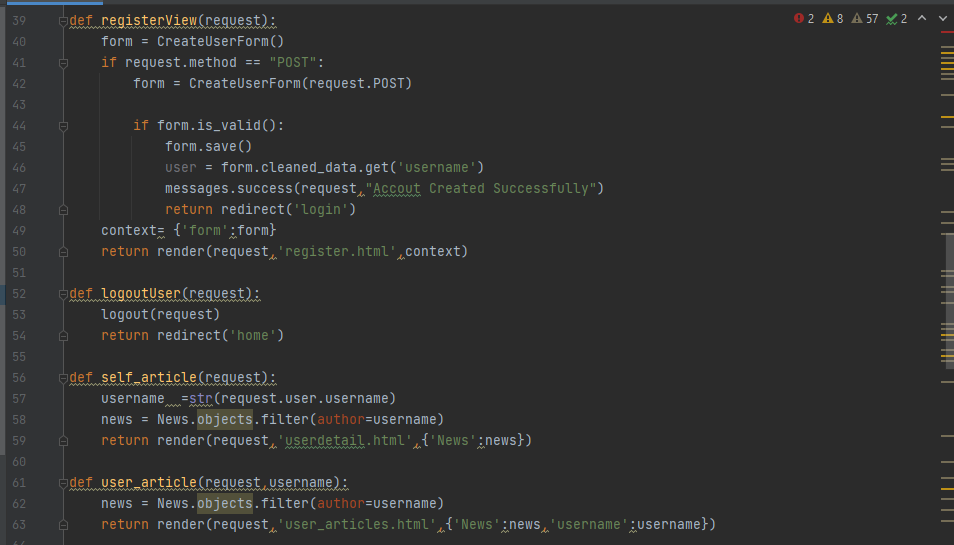
The URL’s for project were defined by the code snippet given below.

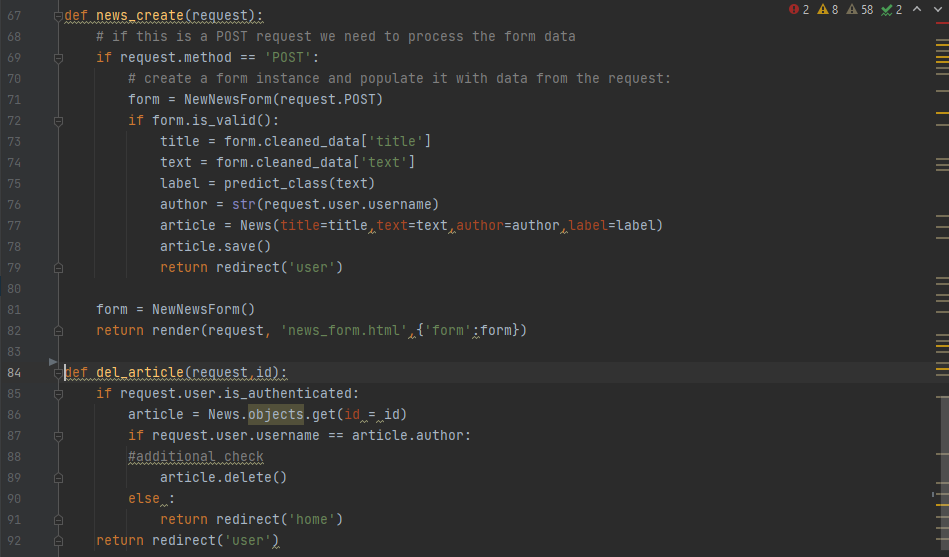
Forms for taking information from author is implemented using the following code snippet.



For every URL in the url.py a view is defined in views.py to render various HTML pages. The code for rendering the required pages is shown below.







**HomeListView** - this view is responsible for redering the home page the screenshot of the webpage is given below. It displays the aggregated news articles.

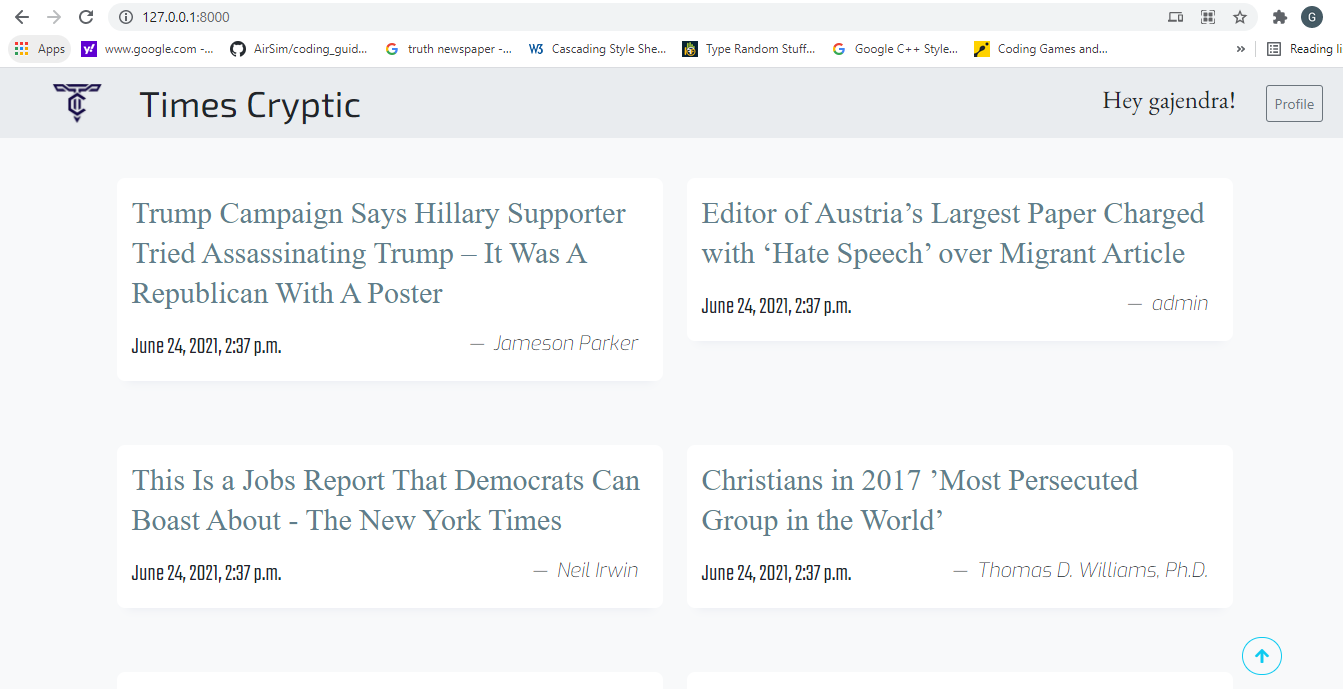


Fig 2.1

**NewsDetailView** - this view is responsible for showing detailed article. The screenshot of webpage is shown below.

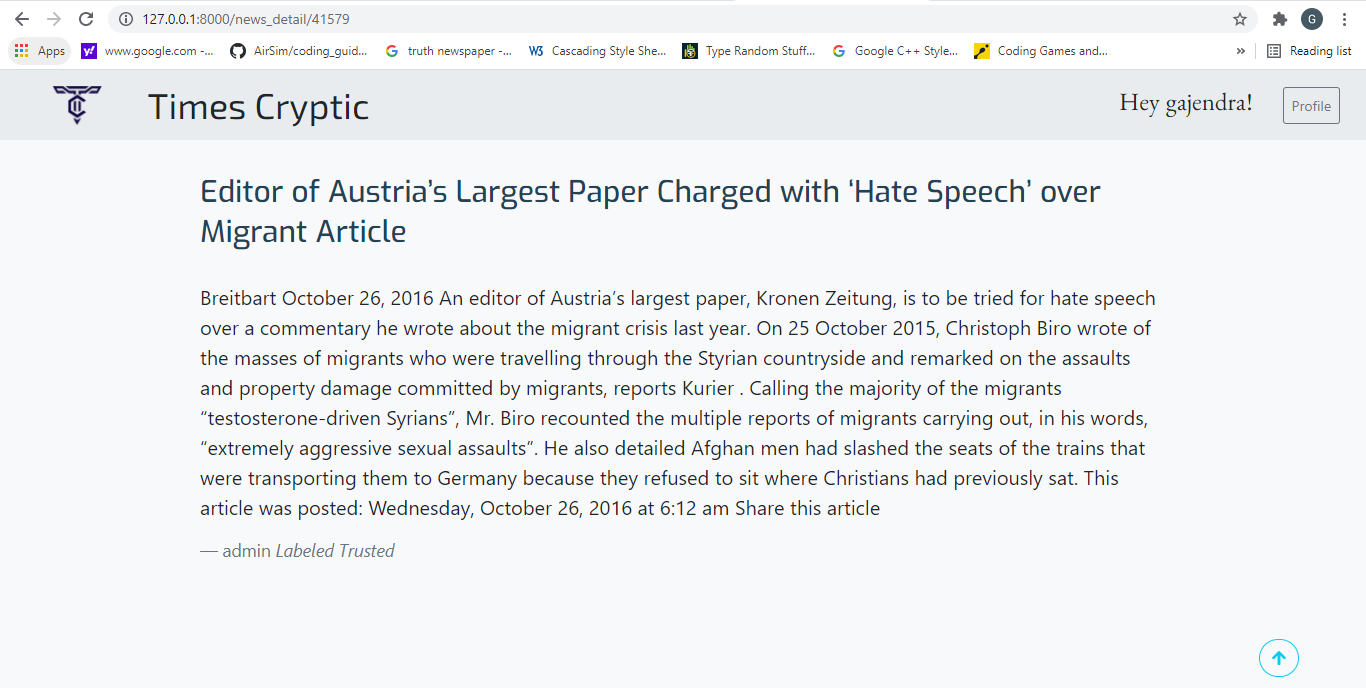


Fig 2.2

**LoginView** - this view is implemented for logging in the authors for viewing and uploading news articles. We used Django admin to implement authentication. The webpage that renders through this view is shown below.

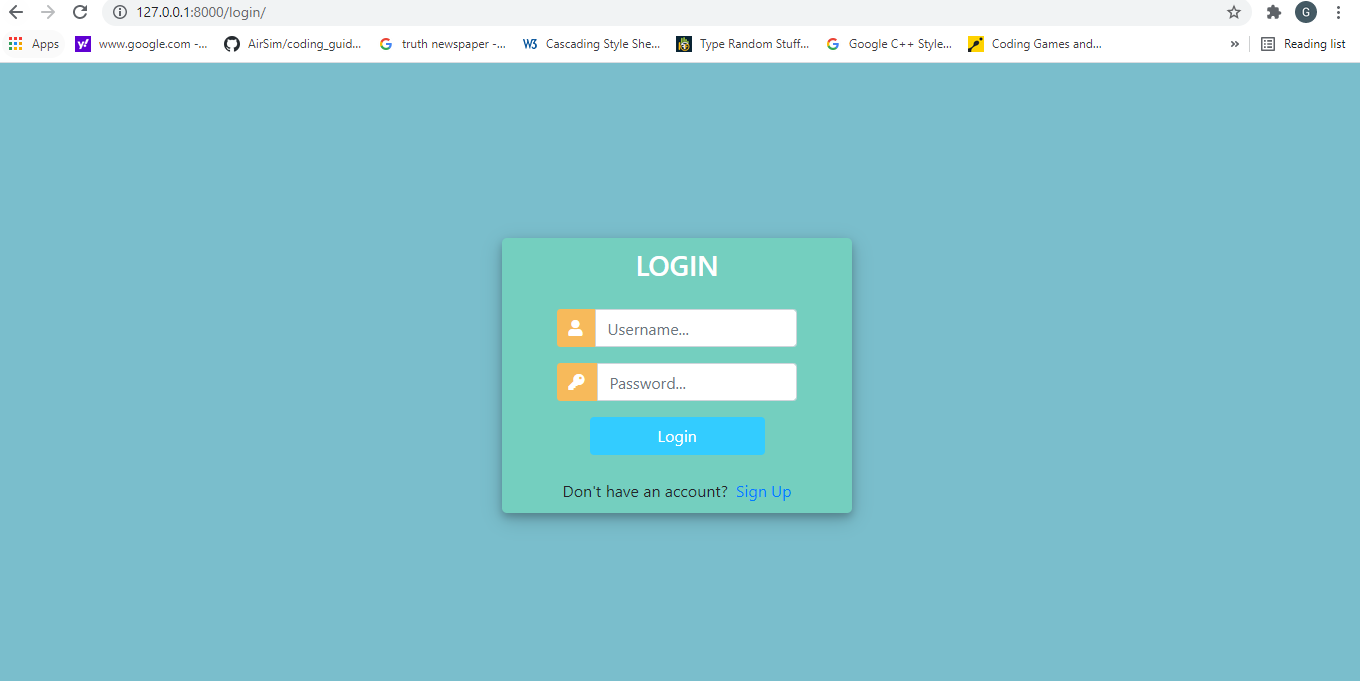


Fig 2.3

**registerView** - this view is implemented for signing up the authors. We used Django admin to implement authentication. The webpage that renders through this view is shown below.

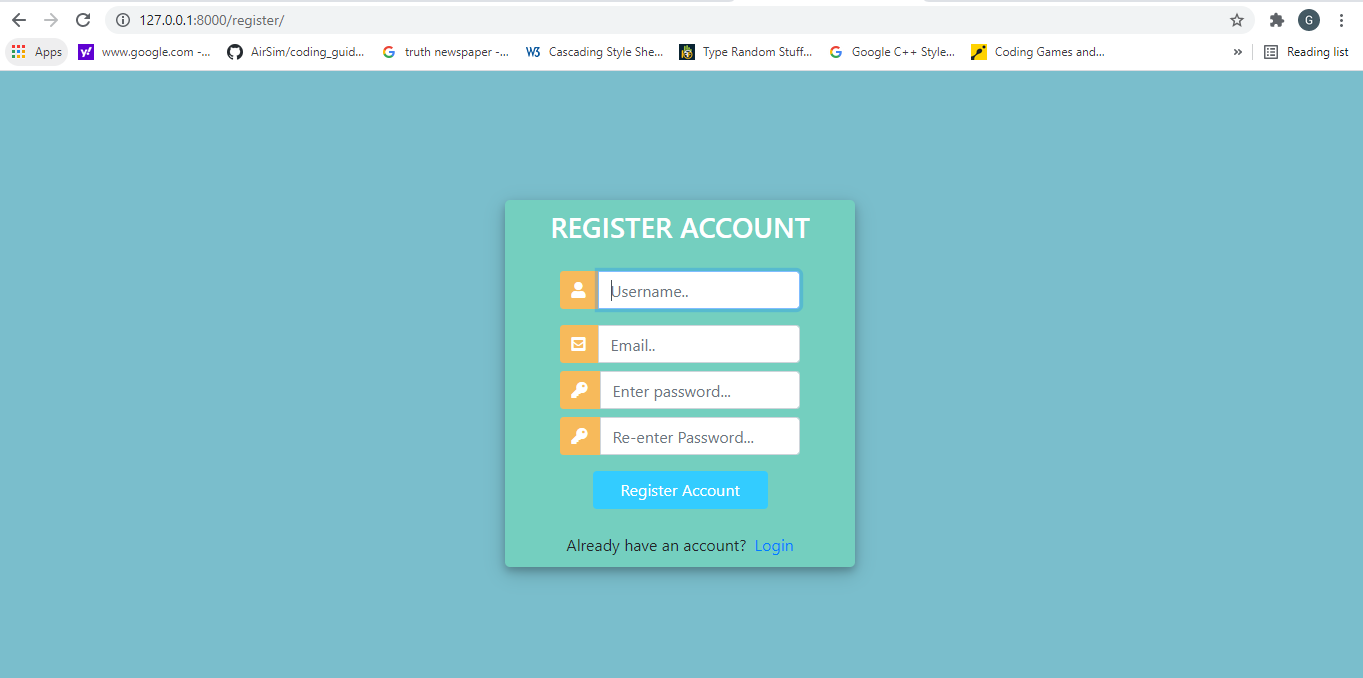


Fig 2.4

**self\_article** - it displays all the news articles that are created by the logged in author. This view gets triggered when user hits profile button on the top right in fig 2.1 or 2.2

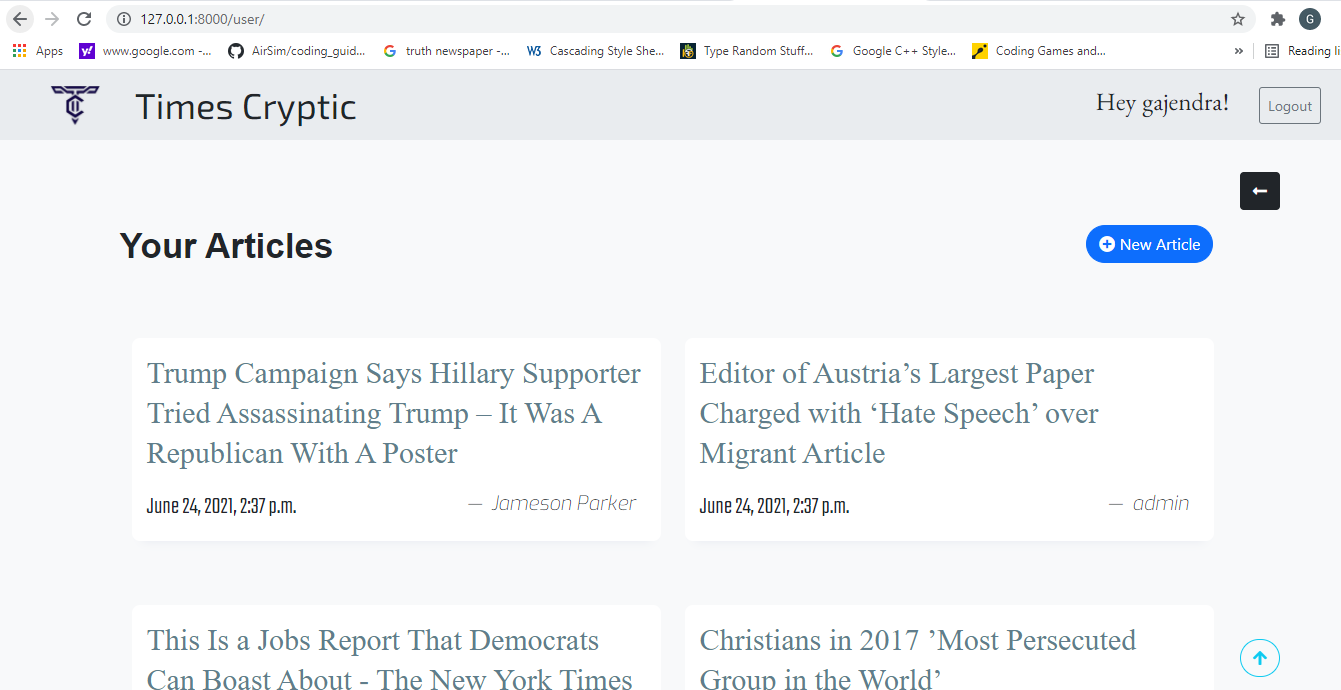


Fig 2.5

**user\_article** - it displays desired authors news articles and has same look as fig 2.5.

**logoutuser** - this view logs out the current logged in user and this view gets triggered when the user clicks on logout button displayed in fig 2.5.

**news\_create** - this view gets triggered when user clicks on add new article button on fig 2.5. it creates a new news article which will get labelled as trusted or not trusted by LSTM model and posted to main homepage.

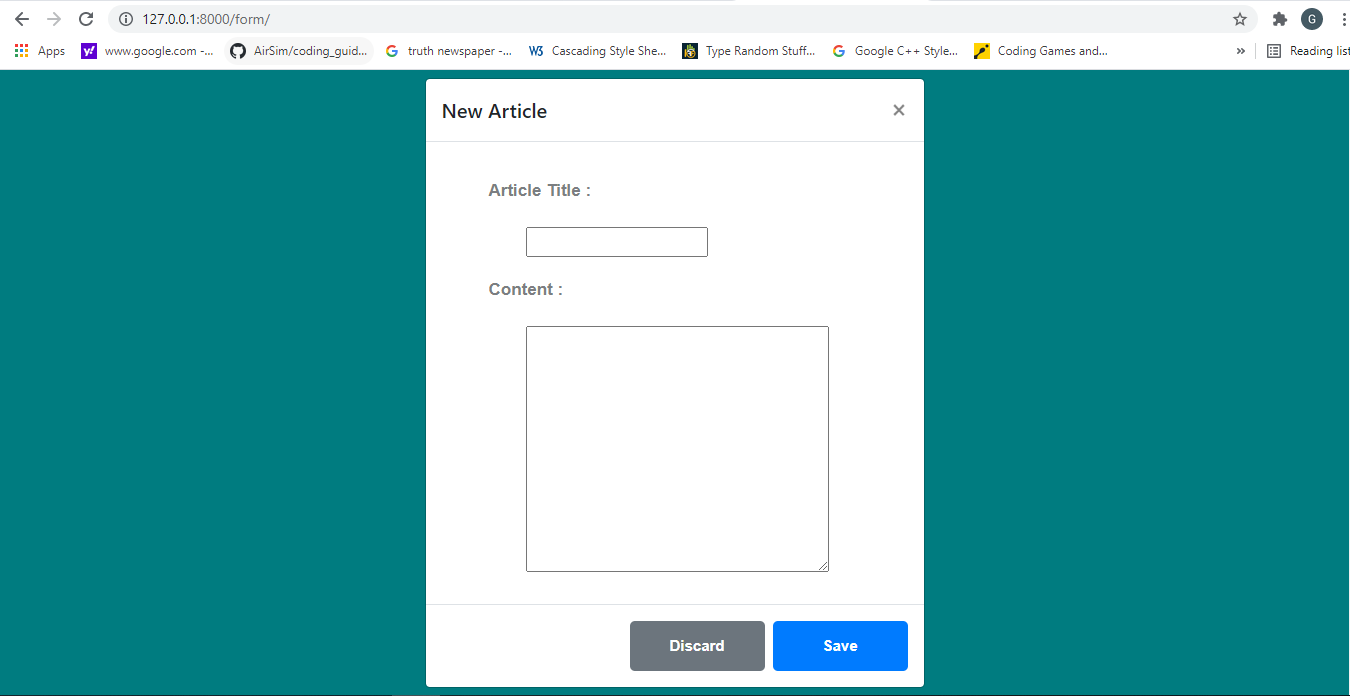


Fig 2.6

**del\_article** - this view is implemented to delete the news articles posted by the authenticated author as per author’s convenience it can be done by button below the detailed news pages in logged in condition.

# CONCLUSION

Manual Classification of News is a very tedious and requires in-depth knowledge of the domain that. In this project, we have developed a Sequence classifier that does this tedious task and gives user the easy of accessibility and reliability. Focus of the project is to give the common man the power to fact-check any news/article and provide a platform to journalists to publish their article and utilize the capabilities of our machine learning model.

Our model yields accuracy of 99.42% on the test dataset which makes it quite reliable as the dataset has more than 20 thousand datapoints. This opens up many possibilities for future uses.

The website that we have developed provides a platform to each of our user to post their article and get tested by our model. This provides user with the power of speech as well as being labelled trusted if they stay true in their words.