MEDIA BIAS MONITORING

S6 Mini Project

Kiran Varghese, Mohammed Farhan, S Vaisakh

June 29, 2016

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)

AIM

To classify an RSS Feed fetched from a media source and classify them using the K Nearest Neighbour Algorithm into appropriate categories.

ABSTRACT

Media Bias monitoring is the activity of monitoring the output of the print, online and broadcast media for identifying its bias towards specific areas like Business, Sports, Politics, Entertainment, Health etc. Enable the theme by loading

INTRODUCTION

- Internet-based resources, such as online newspapers, blogs, and discussion forums,have increased in number, volume, and coverage and its content is regularly viewed and read by lots of people.Lately,the media has started showing bias towards specific areas.
- By the use of the popular classification algorithm, K nearest neighbour algorithm it is possible to classify the article/text fetched from the websites into specific categories.

MEDIA BIAS MONITORING

DESIGN

- In order to classify the document, the first step is to do data cleaning, including stemming and lemmatization, and removing stopwords.
- The un-normalized TF-IDF has been chosen to calculate the vector for each document.
- Each article is then classified using KNN Algorithm and Result is displayed
- The output is displayed on the webpage and stored in a database

KNN CLASSIFICATION

- K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).
- The distant function used in this case is the cosine similarity.
- The KNN algorithm is implemented in python with K=5 using cosine similarity as the method to calculate distance.

KNN CLASSIFICATION - CONTINUED

- Training Phase
  - A number of objects is Classified Manually. This is the training set. The feature vectors and class labels of these samples are stored.
  - The computer reads in this set of objects.

KNN CLASSIFICATION

- 2.Classification phase
  - • A new, unclassified input object (test sample) is classified by a majority vote of its neighbors.
  - – The neighbors are taken from the training set.
  - – Distances from the test sample object to all stored sample objects are calculated, and the k nearest neighbors of the object are selected.
  - – There are different ways to assign a particular class to the object . An object is assigned to the class C, if it is the most frequent class label among the K nearest training samples . If K = 1, then the class of the nearest neighbor is assigned to the object . This special case (K = 1) is called the nearest neighbor algorithm.

CONTROL FLOW DIAGRAM

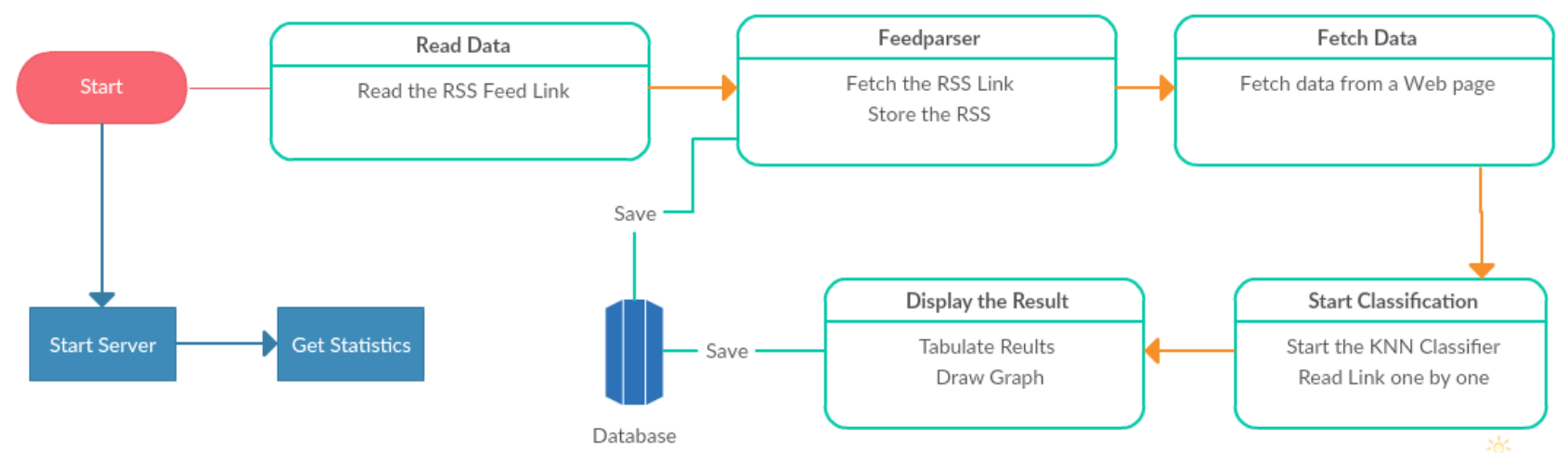

Figure: Control Flow Diagram

FLOWCHART

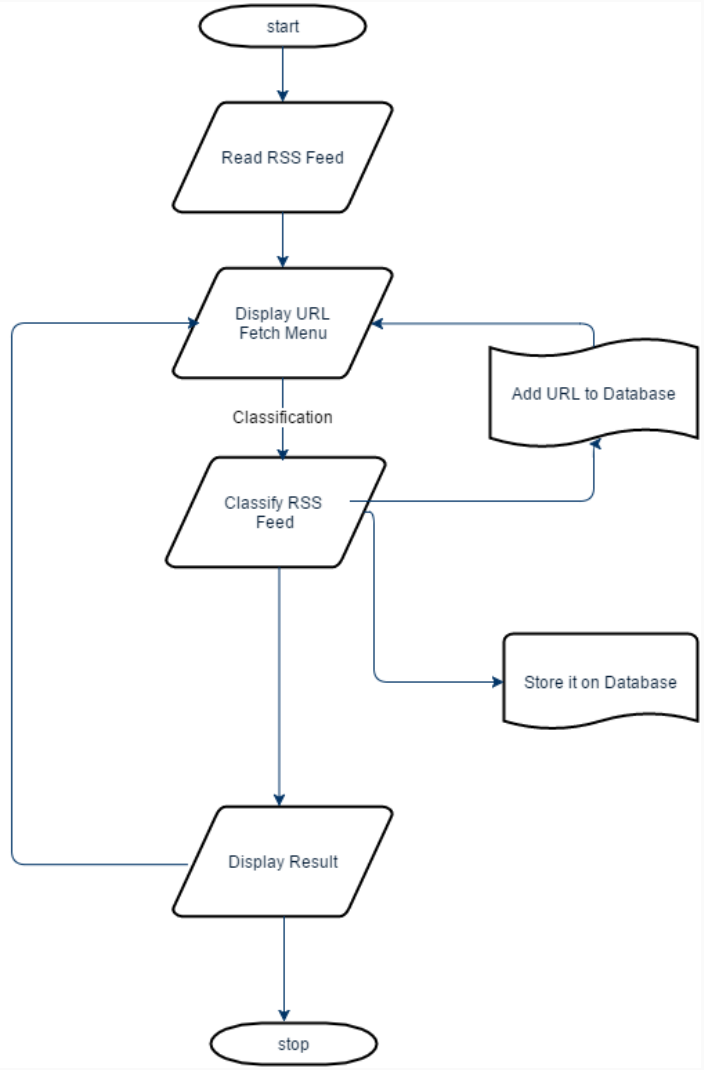

Figure: Flowchart

GENERALISED DIAGRAM

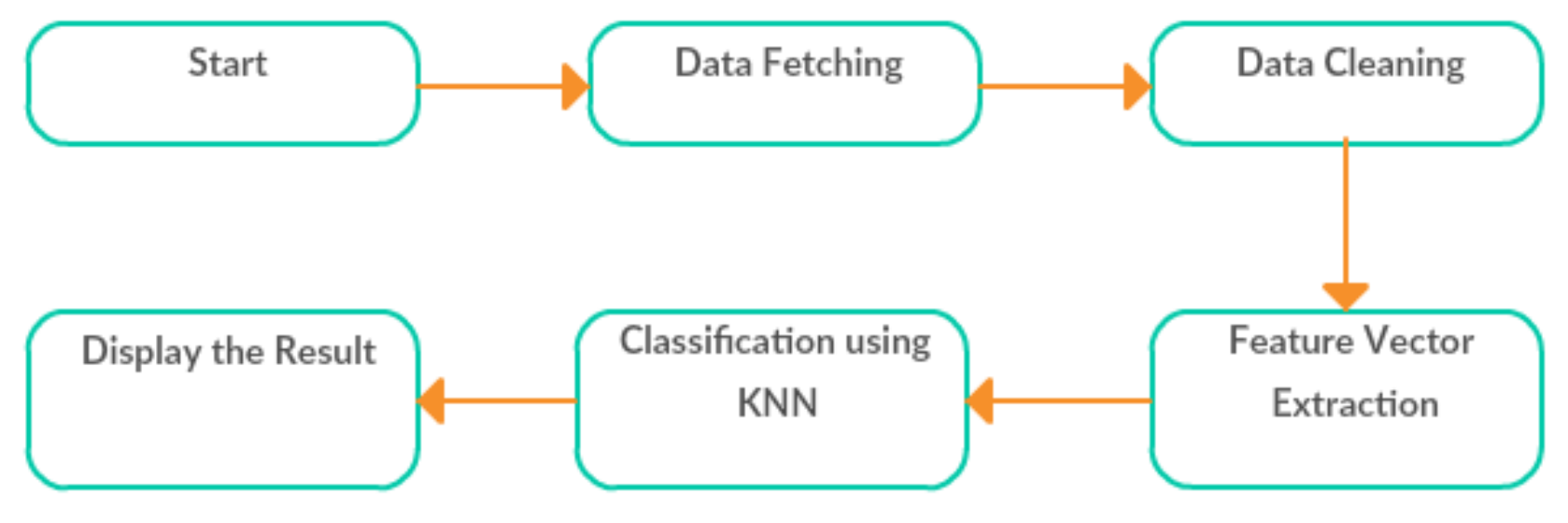

Figure: Generalised Diagram

IMPLEMENTATION

- Python is used for coding.
- An interface is created using Django.
- The applications created are Create, display, Predict, fatch.
- HTML and bootstrap is used for creating templates.
- MySQl is used as the database.

SCREENSHOTS

SCREENSHOTS

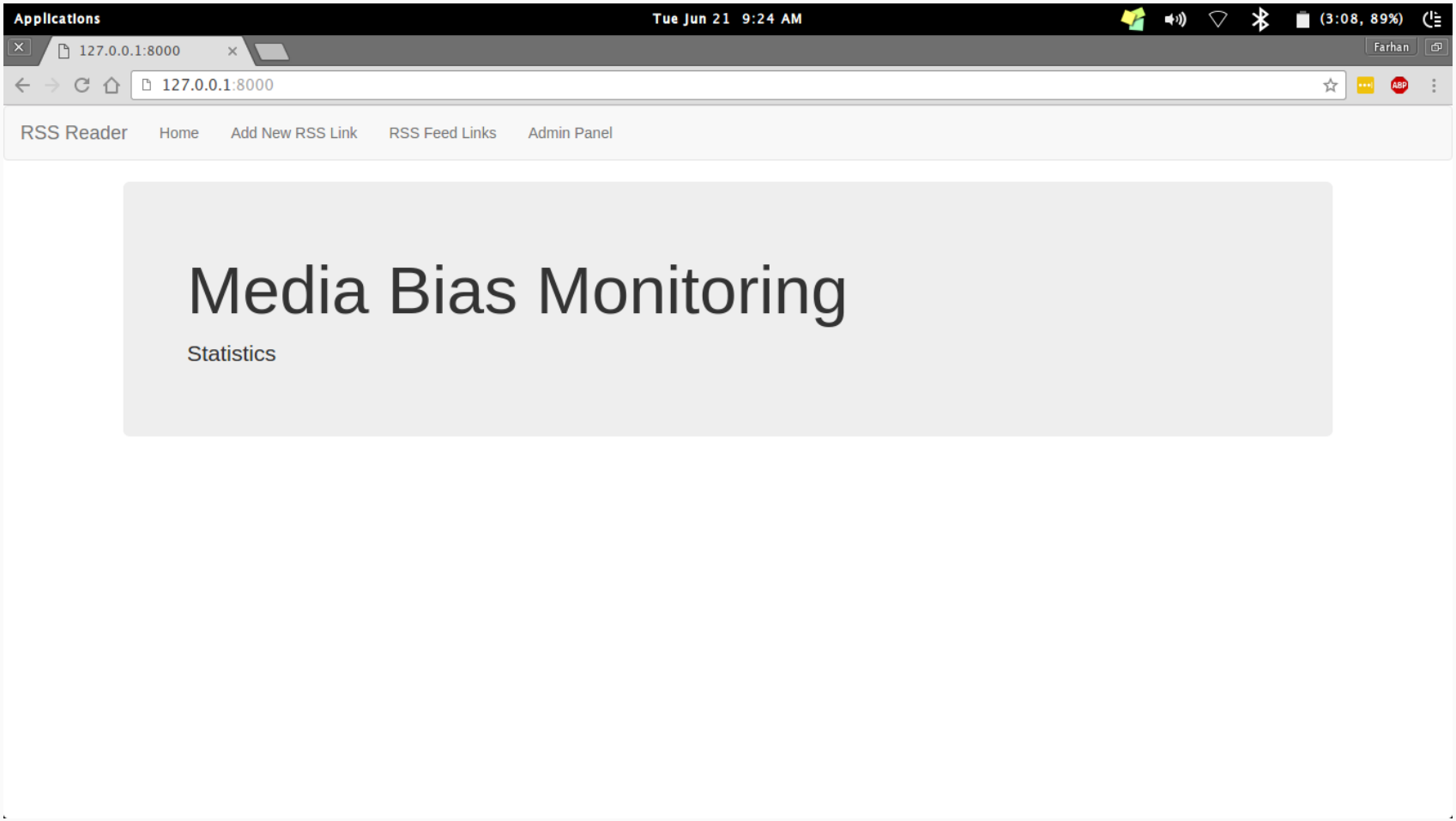

Figure: Homepage

SCREENSHOTS

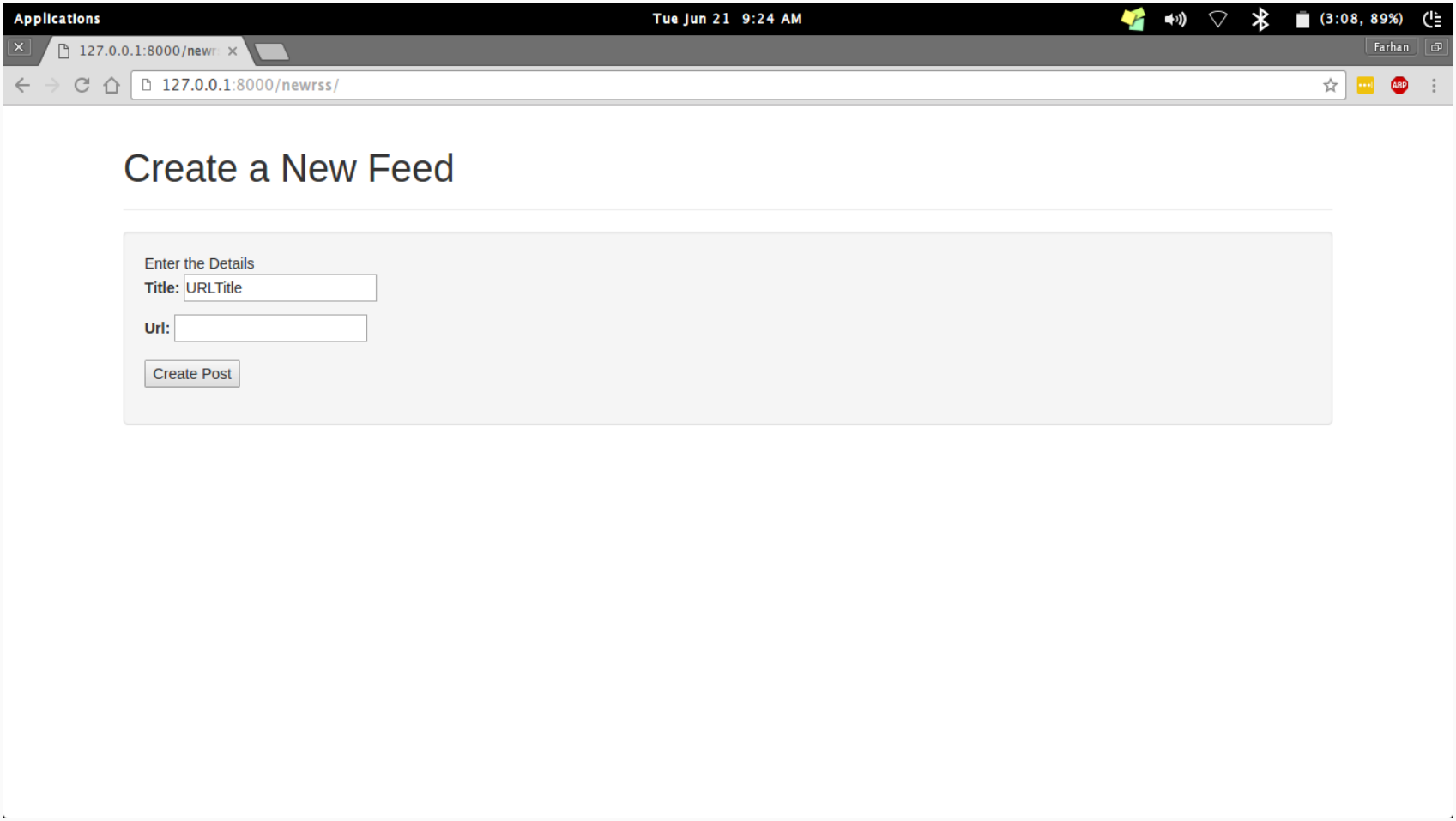

Figure: Create a new Feed

SCREENSHOTS

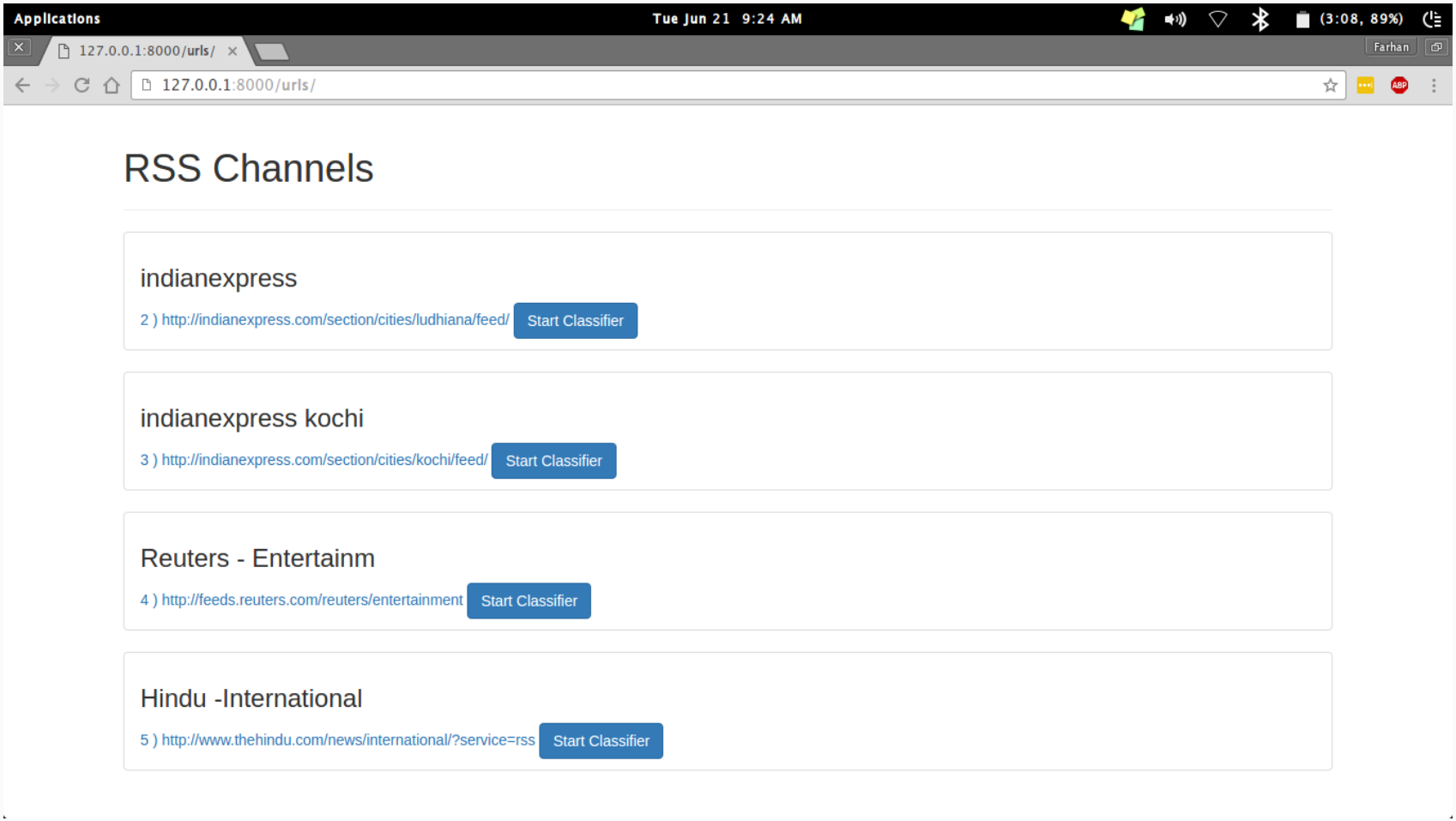

Figure: RSS Channels Display

SCREENSHOTS

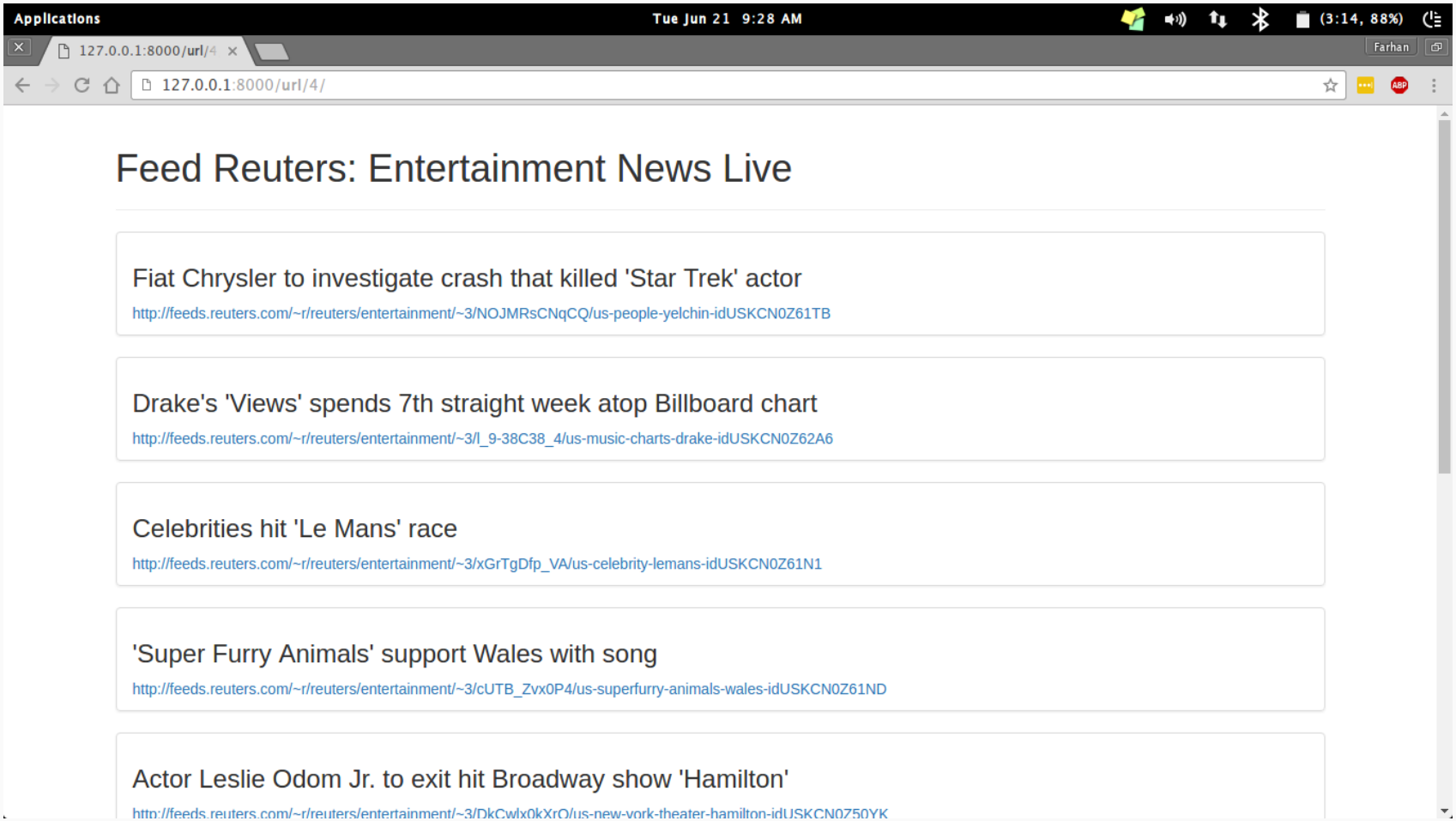

Figure: RSS Feed Display

SCREENSHOTS

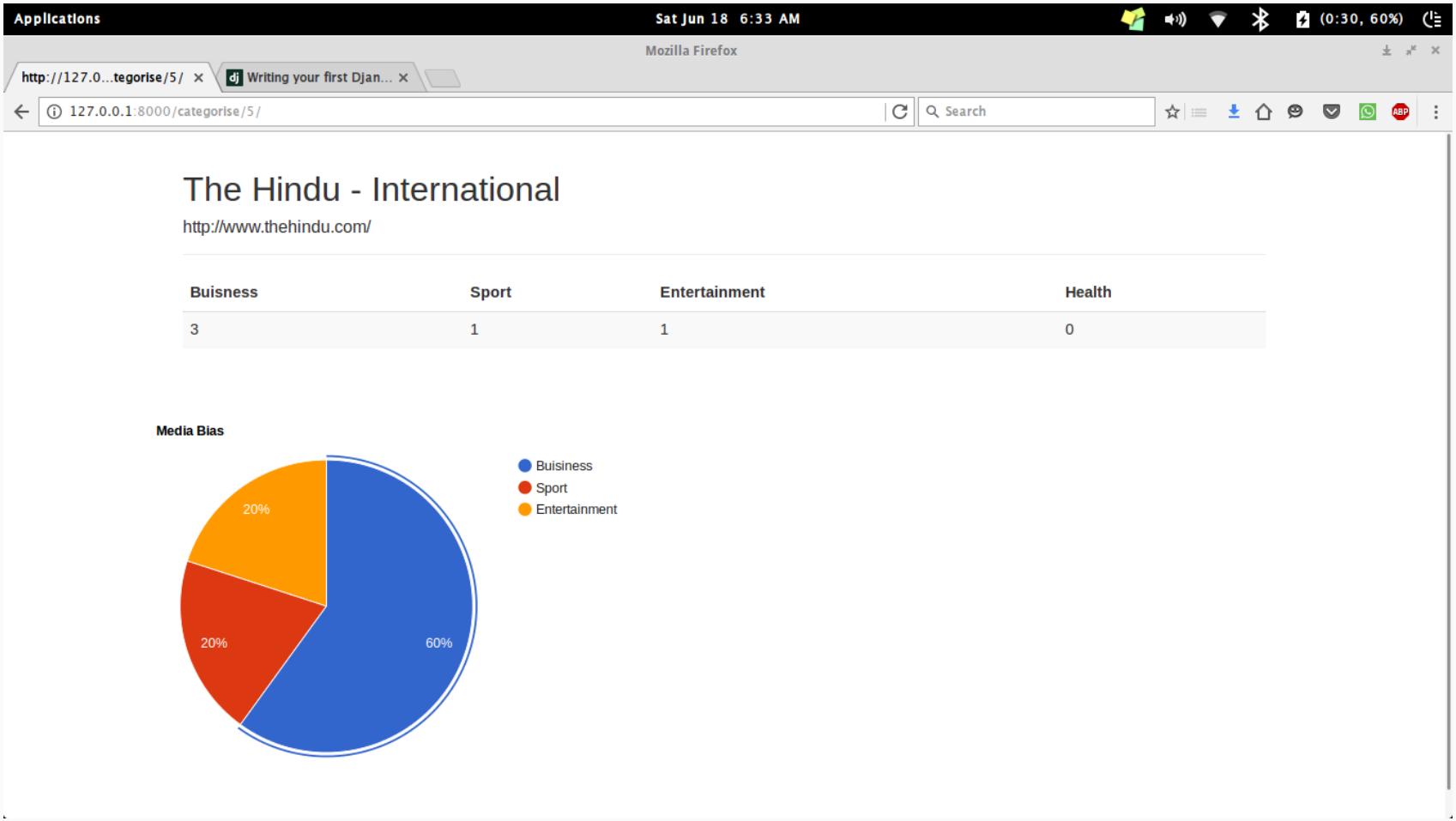

Figure: Result

CONCLUSION

CONCLUSION

Media Bias Monitoring was Implemented using KNN Classifier.Our classifier can aggregate the content in the media feeds to determine the bias towards a particular category and then generate simple data displays and charts, which enable one to get a grasp on current media bias

REFERENCES

K-Nearest Neighbor Text Categorization Method
https://www.usenix.org/legacy/event/sec02/full_papers/liao/liao_html/node4.html

BoilerPipe https://github.com/misja/python-boilerpipe

Django documentation, [online]
https://docs.djangoproject.com/en/1.9/.

Python FeedParser https://pythonhosted.org/feedparser/introduction.html

Thank you