



FactU: An Android Application for News Verification Using Sentence Similarity Analysis and Hidden Markov Model

Emmanuel B. Constantino Jr. and Homer C. Malijan



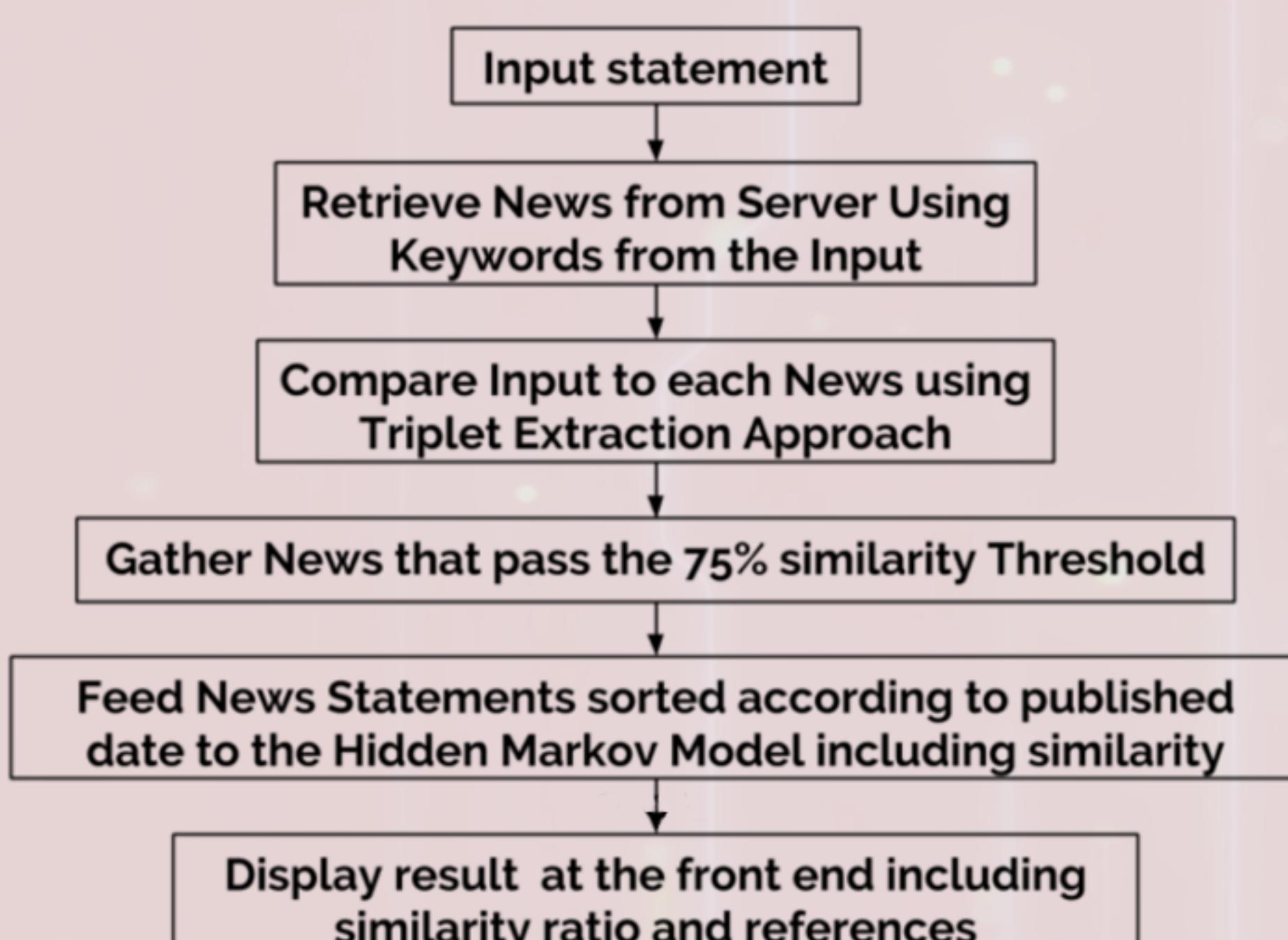
INTRODUCTION

As the literacy rate of the world increases, verification of information will be the next concern. The world is also starting to appreciate the World Wide Web, making it easier for people of any age bracket to surf for information on the Internet. This causes the Internet to be prone to hackers, trolls, or fake online information since most users believe what they see on the Internet.

OBJECTIVES

- To develop an Android application using Python;
- To web crawl various news websites to feed the database with information;
- To evaluate whether or not a news article supports a particular statement using sentence similarity analysis; and
- To assess the effectiveness of using Hidden Markov Models (HMM) to classify news statements.

METHODOLOGY



ABOUT THE AUTHORS

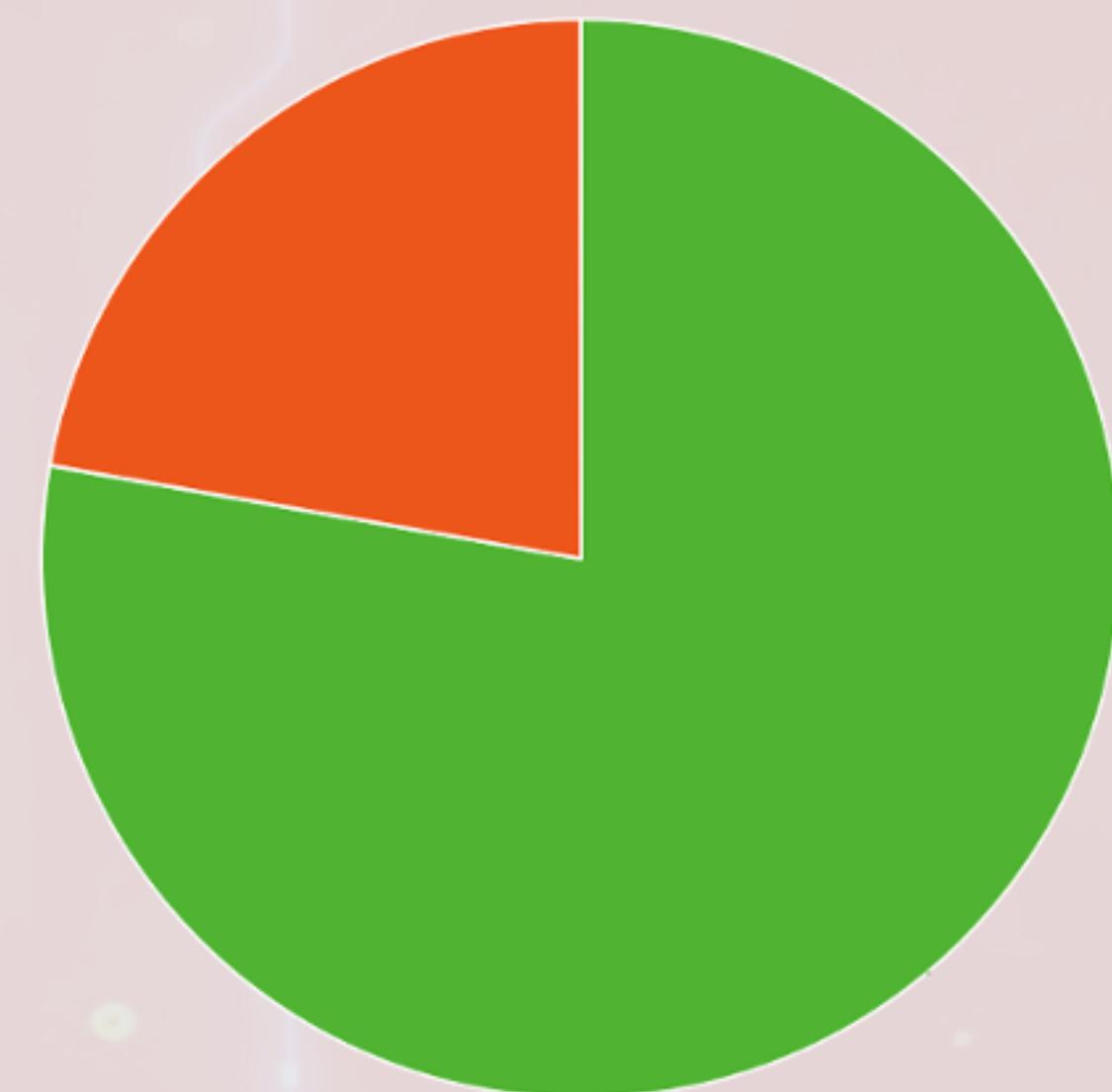


Emmanuel B. Constantino Jr.
BS Computer Science
Batch 2013
Class of 2017



Homer C. Malijan
BS Computer Science
Batch 2013
Class of 2017

RESULTS AND DISCUSSION



■ Success ■ Fail

Data gathered from testing resulted in an 77.78% success rate. Some input resulted in 'Insufficient Evidence' as an output because no entry from the database was able to pass the 75% similarity threshold with the given input, thus, no sense to forward it to the HMM module. On the other hand, some results were erroneous due to having a high similarity percentage with its corresponding database entry but also having a lot of other entries with significant similarities from the other classification making it erroneous. One of the test cases resulted into 'Cannot access link' because for some reason the specific website does not let spiders crawl its documents to protect its information.

CONCLUSION

Out of the three approaches to calculate for a score to classify news statements, the Triplet-extraction approach, was found to produce the most realistic results when compared to the results received from the survey.

Although Hidden Markov Model generated a realistic result, other Artificial Intelligence computational models may be used to garner a more definite result. The results discussed also suggest the need for a bigger database with more information to calculate a better result.