

**OPINION MINING MODEL FOR SINHALA TO ENGLISH
TRANSLATED DATA**

WEERASINGHEGE NUSHAN CHAMARA WEERASINGHE

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Big Data Analytics

School of Computing

**Informatics Institute of Technology, Sri Lanka in Collaboration with Robert
Gordon University, Aberdeen, Scotland**

2022

Abstract

In the world, there are many languages. Every language builds with different Characteristics and Patton. Let us take the significant differences between languages. The first and most common difference is the alphabet. Furthermore, other changes are sounds, word construction, Word meaning, and the writing pattern. Compared to all languages, English is the most spoken language in the world. The effect is that most of the technology started around the English language. Natural language processing is one of the technologies that started using the English Language. Natural language processing is used to create interactions between computers and human language. Most research, support documents, and libraries are built starting and using English. Researching NLP & machine learning with other languages is very rare. However, the use of this kind of technology is getting higher day by day. So minor language users have to find an alternative to getting results somehow. This research will focus on creating an Opinion Mining Model for the Sinhala language.

This project was completed with an opinion mining model using CNN with own labelled dataset. Speciality is this model built with translated data. Google translate is the tool used to translate Sinhala sentences into English. Some time translation is not accurate compared to the Sinhala sentence. This model includes this error also. If I trained the model by correcting translated sentences, we could not directly input the translated sentences for the prediction part. This model is created to get sentiment from the translated sentence without making changes to the sentence.