



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**APPLICATION OF CONFORMAL PREDICTION ON
SEMANTIC SIMILARITY SCORING**

A dissertation by

Shohan Anthony Saverimuttu

Supervised by

Mr. Sudharshan Welihinda

Submitted in partial fulfillment of the requirements for the BSc in Computer Science
degree at the University of Westminster.

July 2023

ABSTRACT

Even with the proliferation of machine learning models and techniques and the use of predictive software for corporate or research and academic purposes, it is understood that there lingers a lack of transparency and risk that restricts the adoption of such machine learning techniques. How well can you trust the output result of a model? How reliable is it? These are questions observed by the author after thorough review of existing literature and academic papers.

This call for the application or tool that will incorporate the indication of the level of confidence in a result or output brought about by the introduction of Conformal Prediction that uses the notion of algorithmic randomness to make reliable predictions on the confidence level for an individual example. Since predictive software varies from domain to domain and is vastly applied across many problems, the author has decided to apply conformal prediction onto the semantic similarity score. For this report, the author displays an MVP for a semantic similarity score system that has conformal prediction incorporated into it.

The author was able to incorporate the technique of conformal prediction to produce a conformal interval to help improve the reliability of outputs. The semantic similarity model utilized for the thesis has a Pearson correlation coefficient of 0.85 out of 1 which promises high accuracy. The conformal predictor produces a prediction interval with an error coverage of 0.05.

Keywords: Predictive software, conformal prediction, semantic similarity, Sentence BERT, Inductive conformal prediction, NLP.