**Product Search Improvement in Amazon via Natural Language Processing**

Github link: (<https://github.com/BIJESHPATEL369/CSCE-5290-Natural-Language-Processing-Project> )

**Team Members**

Shreeti Upreti ([upretiupreti@my.unt.edu](mailto:upretiupreti@my.unt.edu))

Bijesh Patel Vachanni ([bijeshpatelvachanni@my.unt.edu](mailto:bijeshpatelvachanni@my.unt.edu))

**1.** **Goals and Objectives:**

● **Motivation**

The motivation to work on this project has come from an interest to work on solving real-life problems. The Amazon KDD Cup Challenge brings forth the opportunity to work on tasks implemented in the e-commerce website to enhance its predictive performance for product searches in an intuitive manner. The challenge does not only provide the dataset generated from the Amazon website but also guides the participants through the logical approach one can take to talk about problems of this nature. Moreover, the project aligns with the concepts like text classification, minimum edit distance calculation, e.t.c, that are covered in the [CSCE 5290 Section 001 - Natural Language Processing (Spring 2022 1)](https://unt.instructure.com/courses/62420) class.

● **Significance**

The Amazon KDD Cup Challenge as a project will be a platform to apply Natural Language processing concepts to practical usage. The challenge has three sub-tasks, each dealing with different problems. Of the diverse tasks listed for the challenge, the tasks that best fit the possibility of applying the theoretical aspects learned thus far will be chosen and worked upon. This project provides the prospect of further leveraging the NLP concepts into practice. Given the time and resource constraints, this project will cover two out of the three tasks in the challenge.

**● Objectives**

The objective is to find a better approach to information retrieval on semantic matching with user queries and relevant products. The products will be ranked based on the most to the least relevant products in the first task, given each Query as input. The second task will classify the resultant products from the queries into four categories: exact, substitute, complement, and irrelevant. This ranking approach can leverage E-commerce to find the most profitable product ranking system to facilitate customers' search queries.

**● Features**

Task 1:

The input will contain the user query, and the output will be products for each Query that will be ranked based on the most relevant ones.

Task 2:

The input will be product and Query pair along with product features like product title, description, brand, colour, locale, etc., and the target labels will be following four categories:

* Exact: the most relevant product from the Query.
* Substitute: somewhat suitable product from the Query.
* Complement: the product will be mapped with other products to match the Query.
* Irrelevant: Product that is not relevant to the Query.

**2.** **References**

1. <https://www.aicrowd.com/challenges/esci-challenge-for-improving-product-search?source=mlcontests>
2. <https://towardsdatascience.com/bert-for-dummies-step-by-step-tutorial-fb90890ffe03>
3. <https://www.analyticsvidhya.com/blog/2020/08/information-retrieval-using-word2vec-based-vector-space-model/>