**Predicting Yelp Rating Polarity:**

**Leveraging NLP for Sentiment Analysis of Local Business Reviews**

Team 9 – Indranil Pal, Yasser Parambathkandy

**George Mason University**

**AIT526 Natural Language Processing**

**06/26/2023**

**Predicting Yelp Rating Polarity**

**The goal of this project is to estimate the polarity (0 or 1) of Yelp ratings for local business reviews based on the review text. This will be accomplished by leveraging Natural Language Processing (NLP) models with a focus on sentiment analysis. Yelp, a popular review-based website, provides a platform for people to discover and exchange information about local businesses. Often, when consumers conduct quick searches for businesses, they tend to rely solely on the star rating without delving into the review language. As a result, this project aims to explore potential connections between the wording of reviews and the corresponding ratings. The dataset for this project has been obtained from Kaggle and is derived from the Yelp Dataset Challenge 2015. The Yelp reviews polarity dataset has been modified to classify stars 1 and 2 as negative (label: 0) and stars 3 and 4 as positive (label: 1). Therefore, the dataset represents a binary classification task. It consists of over 560,000 training samples and 38,000 testing samples, with an equal distribution of samples for both labels. The feature column in the dataset corresponds to the review text, while the target column represents the label.**

**In this project, various NLP algorithms and Deep learning techniques such as Recurrent Neural Networks (LSTM), will be investigated and compared. The objective is to select a suitable algorithm that can effectively predict the polarity of Yelp ratings. The models will be implemented using PyTorch, an open-source framework known for its efficient utilization of GPUs.**