Natural Language Processing
Final Project Guidelines and Suggestions

Rules:

- All projects are **individual** projects. No paired or team projects.
- Please **see Canvas for due dates** for the various deliverables.
- Each student will submit a document on Canvas describing their **project proposal**.
    - Some project suggestions are below. Some variation from those project descriptions is possible.
    - If you want to do something different than any of those, please pre-check with me before you submit your proposal. Essentially, we want to make sure that the scope of the project is something that can be reasonably done in that time available.
- Each student will submit a **Related Work** paper, 2-3 pages, that describes prior research on the topic they have selected.
- Each student will submit a **Progress Report**. That progress report should include
    - a description of initial code development, and
    - screenshots of what is working.
- Each student will submit a 10-minute **Recorded Presentation** of their project.
- Each student will do a **Peer Review** of three (3) other student projects by watching their recorded presentations.
- The **Final Project Report plus all code** will be submitted at the end of the semester.
    - Each project submission must include a README file and any other instructions that would be necessary for me to run their code.

Project Name	Description	Evaluation
Sentiment Analysis on Amazon product reviews	Build several different classifiers to determine whether reviews are positive, negative, neutral or **other rating (e.g., # of stars)**	Use accuracy, precision, and recall on a labeled dataset.
Named Entity Recognition (NER) for News Articles	Implement at least two different methodologies for an NER system to identify entities in news articles.	Compare performance with standard NER benchmarks (e.g., F1 score).

Text Summarization Using Transformer Models	Automatically summarize articles using two different pre-trained transformer models.	Use ROUGE scores to compare generated summaries to human-written summaries.
Toxic Comment Filtering	Train a classifier using word embeddings (BERT, Word2Vec) and CNNs/LSTMs.	Use accuracy, precision, and F1 score on a labeled dataset.

Project Name	Description	Evaluation
Poetry Generation with GPT-2 or other	Build two different language generation systems to generate poetry based on user prompts.	Evaluate coherence, creativity, and fluency through human evaluation.
Machine Translation between English and Spanish	Implement two different systems to translate English to Spanish.	Use BLEU scores to measure translation accuracy.
Misinformation Detection in Social Media Posts	Fine-tune GPT-3/BERT to detect false claims, misleading info, or conspiracy theories.	Use accuracy and F1 scores on a labeled dataset like FakeNewsNet.
Part-of-Speech (POS) Tagging for Social Media Posts	Implement two different methodologies for POS tagger for informal social media text.	Compare the model's output with a tagged corpus, measuring accuracy.

Project Name	Description	Evaluation
Text Generation for Autocomplete	Develop a system that suggests word completions as users type.	Measure keystroke savings and prediction accuracy.

Sarcasm Detection in Tweets	Implement two different methodologies to detect sarcasm in tweets.	Evaluate performance using F1 score on a labeled dataset.
Topic Modeling for News Articles	Implement topic modeling using LDA to identify topics in news articles.	Evaluate topic coherence through manual inspection and metrics.
Author Attribution Using Stylometry	Use two different methodologies to predict the author of a text based on writing style. [I had one student analyze rap lyrics to determine the artist.]	Measure accuracy on a dataset of texts from multiple authors.

Project Name	Description	Evaluation
Emotion Detection in Text	Implement a system to detect emotions in text based on semantic content. Go beyond basic sentiment (positive/negative).	Evaluate using accuracy and F1 scores on an emotion-labeled dataset.
Question Answering System Using a Knowledge Base	Compare 2 or more diffierent QA system that retrieves answers from a knowledge base.	Measure accuracy by comparing predicted answers with correct ones in a benchmark dataset.

Plagiarism Detection Using Sentence Similarity	Create a system to detect plagiarism by comparing sentence similarity. Create multiple models and compare their performance.	Measure precision, recall, and F1 score on a labeled dataset of original and plagiarized content.

Some Sample Proposals

Sample Proposal #1

Project Proposal: Misinformation Detection in Social Media Posts

The spread of misinformation on social media is a major challenge in the digital age. This project aims to fine-tune **BERT** to detect **false claims, misleading information, and conspiracy theories** in social media posts. Using a labeled dataset such as **FakeNewsNet**, the model will be trained to classify posts as either **misinformation or factual** based on linguistic patterns. The goal is to explore whether BERT can effectively distinguish misinformation from reliable content, improving upon traditional text classification methods.

To evaluate the model's effectiveness, we will measure **accuracy** and **F1-score**, ensuring robust performance across different types of misinformation. The project will also include an **error analysis** to identify common failure cases and discuss potential improvements. This work will provide insight into the **strengths and limitations of transformer-based models in misinformation detection**, with implications for fact-checking and automated content moderation.

Sample Proposal #2

Project Proposal: Topic Modeling for News Articles

Understanding the main themes in large collections of news articles is essential for organizing information and identifying trends. This project will implement **topic modeling** to automatically discover topics in a dataset of news articles. Techniques such as **Latent Dirichlet Allocation (LDA)** or **Non-Negative Matrix Factorization (NMF)** will be used to extract meaningful topics based on word distributions. Additionally, more modern approaches like **BERTopic**, which leverages transformer embeddings for improved topic representation, may be explored.

The evaluation will include both **quantitative and qualitative measures**. **Topic coherence metrics** such as **UMass and NPMI** will be used to assess how well the topics capture meaningful word groupings, while **manual inspection** will help determine interpretability. The final analysis

will compare different techniques, discussing their strengths and weaknesses in uncovering coherent and distinct topics in news content.