Nav Sanya

* <https://www.sciencedirect.com/science/article/abs/pii/B9780128190432000034>
* <https://www.sciencedirect.com/science/article/abs/pii/S0933365715000299>
* <https://aclanthology.org/L18-1442.pdf>
* <https://link.springer.com/book/10.1007/978-3-319-68468-0>
* <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01347-1>
* <https://medinform.jmir.org/2020/1/e16023>
* <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2728368.pdf>
* <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6839297.pdf>

Project ideas: <https://neptune.ai/blog/10-nlp-projects>

<https://www.knowledgehut.com/blog/data-science/nlp-projects#top-12-natural-language-processing-(nlp)-projects-with-source-code%C2%A0>

Sebastian

* Papers:

* Project Ideas:

Sales Agent for Saas Products / Customer Relationship Management

* (Onboarding, Email Marketing)

Local Personal Assistant with an open source model

Anusha:



Shreyas

1. **Text Summarization:**  
   Project Idea: Abstractive Summarization using Transformer Models

Description:

Implement an abstractive text summarization system using state-of-the-art transformer models like BERT or T5. The goal is to generate concise and coherent summaries that capture the essential information from longer documents.

Research Papers:

* [Text summarization with pre-trained encoders](https://arxiv.org/pdf/1908.08345.pdf)
* [Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer](https://arxiv.org/abs/1910.10683)

Dataset:

Use the CNN/Daily Mail dataset for abstractive summarization. It contains news articles paired with multi-sentence summaries, making it suitable for training and evaluating summarization models. <https://huggingface.co/datasets/cnn_dailymail>

1. **Multimodal NLP:**Project Idea: Image Captioning with textual sentiment analysis

Create a system that generates captions for images while considering the sentiment expressed in the associated text. The goal is to produce image captions that align with the sentiment conveyed in the textual description.

Research Papers:

* [Show and Tell: A Neural Image Caption Generator](https://arxiv.org/abs/1411.4555)
* Concept-oriented [transformers for visual sentiment analysis](https://dl.acm.org/doi/abs/10.1145/3539597.3570437)

Dataset:

MS COCO dataset for image captioning and the Affective Image Computing (AIC) dataset for sentiment analysis. The AIC dataset includes images labeled with sentiment scores. <https://paperswithcode.com/dataset/coco-captions> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9002643/>

1. **Sentiment Analysis:**

Project Idea: Fine-grained Sentiment Analysis in Product Reviews

Description:

Develop a sentiment analysis model that goes beyond binary sentiment classification and performs fine-grained sentiment analysis on product reviews. Classify sentiments into multiple categories (e.g., positive, neutral, negative) and extract key aspects influencing sentiment.

Research Papers:

* [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805)
* [Aspect-Based Sentiment Analysis with Gated Convolutional Networks](https://arxiv.org/abs/1805.07043)

Dataset:

SemEval 2014 datase. It contains product reviews labeled with sentiment polarities for specific aspects, allowing for a more detailed sentiment analysis. <https://paperswithcode.com/dataset/semeval-2014-task-4-sub-task-2>