Capstone Project Ideas on Natural Processing Language

1. **Topic Modelling**: It is a process to automatically identify topics present in a text object and to derive hidden patterns exhibited by a text corpus assisting to make better decisions. Generally it is achieved using the **Latent Dirichlet Allocation model technique**. The dataset used will be **spooky author identification** dataset sourced from Kaggle .it contains text from works of fiction written by spooky authors of the public domain: Edgar Allan Poe, HP Lovecraft and Mary Shelley.
2. **Sentiment Analysis**: It is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service. We will be using IMDB 50k reviews labelled dataset from Kaggle to build model that predicts the sentiment (positive or negative) from the text review.
3. **Multi-Label Classification**: Make a multi-label classification system that automatically assigns tags for questions posted on a forum such as Stack Overflow or Quora. The data set used will be Stack Lite dataset from Kaggle. The objective will be to predict tag from the question text.