NLP'24 Projects

ID	Title	Description	Dataset
1	topic modeling	Topic modeling is a method for finding a group of words (i.e. topics) from a collection of documents that best represents the information in the collection of text documents. It can also be thought of as a form of text mining - a way to obtain recurring patterns of words in textual data. The topics identified are crucial data points in helping the business figure out where to put their efforts in improving their product or services.	Topic Modeling English Topic Modeling Arabic
2	Fake News Detection	Fake news spreads like a wildfire and this is a big issue in this era. You can learn how to distinguish fake news from a real one. You can use supervised learning to implement a model like this.	Fake News
3	News Group classification	Build a model to classify news data into various categories through text classification. The overall process of model training and testing should be recorded and displayed using a proper chart. The classification results for the best model should be displayed too.	Newsgroups
4	Sentiment Analysis of movie reviews	Design a model to apply Sentiment Analysis of movie reviews using sentiment polarity datasets (Use a machine learning model to classify the movies reviews into positive and negative, train and test model and generate accuracy). The overall process of model training and testing should be recorded and displayed using a proper chart. The classification results for the best model should be displayed too.	movie-review-data
5	Automatic Questions Tagging System	Sites that are specifically designed to have questions and answers for their users like Quora and Stackoverflow often request their users to submit five words along with the question so that they can be categorized easily. But sometimes users provide wrong tags which make it difficult for other users to navigate through. Thus, they require an automatic question tagging system that can automatically identify correct and relevant tags for a question submitted by the user.	stacksample