**7 NLP Project Ideas to Enhance Your NLP Skills**

[[](https://davis-david.medium.com/?source=post_page-----19c8e81c4c58--------------------------------)](https://davis-david.medium.com/?source=post_page-----19c8e81c4c58--------------------------------)

[[Python in Plain English](https://python.plainenglish.io/?source=post_page-----19c8e81c4c58--------------------------------)](https://python.plainenglish.io/?source=post_page-----19c8e81c4c58--------------------------------)

[Davis David](https://davis-david.medium.com/?source=post_page-----19c8e81c4c58--------------------------------)

·

Follow

Published in

[Python in Plain English](https://python.plainenglish.io/?source=post_page-----19c8e81c4c58--------------------------------)

·

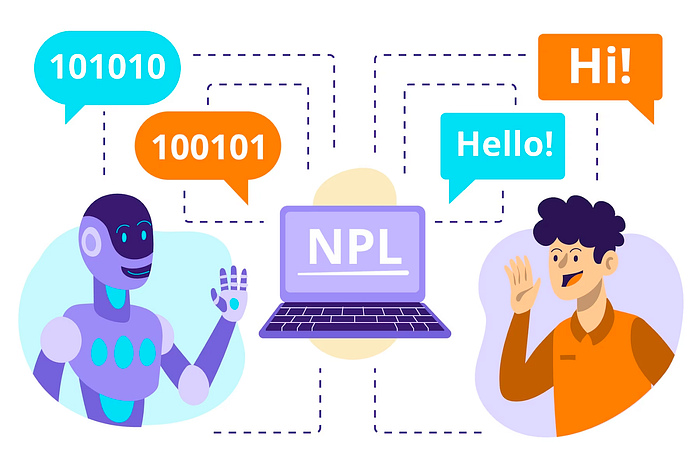
11 min read

·

Sep 19

154

4



from [freepik.com](https://www.freepik.com/free-vector/natural-language-processing-illustration_21743714.htm#query=natural%20language%20processing&position=0&from_view=search&track=country_rows_v2)

Natural Language Processing (NLP) has emerged as a transformative force that reshapes how we interact with information and communicate with machines around the world. NLP is a field at the intersection of computer science, linguistics, and artificial intelligence, focusing on enabling computers to understand, interpret, and generate human language in a way that mirrors human cognition.

As NLP continues to advance in various industries such as healthcare, finance, customer service, and others around the world, it is important to gain practical experience through hands-on NLP projects to become a good Data scientist or NLP engineer.

The scope of NLP applications is both vast and diverse, covering a wide range of industries and use cases. From sentiment analysis and chatbots to language translation, speech recognition, and information retrieval. NLP-driven applications enhance search engines’ accuracy, automate customer interactions, facilitate multilingual communication, and even assist in legal document analysis. This variety highlights how flexible NLP is and emphasizes the urgent requirement for experts who are knowledgeable in its techniques.

Textbook knowledge and theoretical understanding are still valuable components of learning NLP, but they can only take an individual so far. The true mastery of NLP comes from hands-on experience, where learners engage themselves in real-world projects to tackle challenges, experiment with various algorithms, and overcome practical challenges. This will help you to gain invaluable insights about preprocessing text data, feature engineering, selecting appropriate models, fine-tuning parameters, and effectively evaluating results.

In this article, you will learn different NLP project ideas that focus on practical implementation to help you master the NLP techniques and be able to solve different challenges.

**Sentiment Analysis**



[from freepik.com](https://www.freepik.com/free-vector/emoji-satisfaction-meter-small_44156211.htm#query=sentiment&position=0&from_view=search&track=country_rows_v2)

Sentiment analysis is an NLP technique that involves determining the sentiment or emotional tone behind a piece of text, such as a review, tweet, or customer feedback. The main goal of sentiment analysis is to categorize the sentiment expressed in the text as positive, negative, or neutral. This process is important for understanding public opinion, making informed business decisions, monitoring brand reputation, and evaluating customer satisfaction.

In the era of social media and online reviews, sentiment analysis helps businesses understand customer feedback at scale, enabling them to identify areas of improvement and improve customer experience. It assists in monitoring and managing brand perception, as well as predicting market trends based on sentiment shifts.

Several datasets are available for training and evaluating sentiment analysis models. These datasets are often labeled with sentiment labels (positive, negative, neutral) to facilitate supervised machine learning. Some popular datasets include:

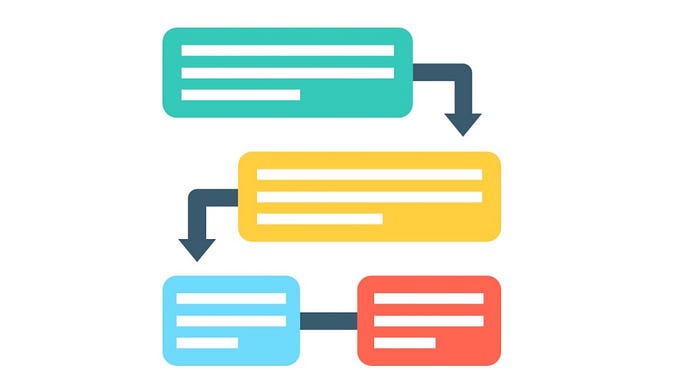
* **IMDb Movie Reviews:** A dataset containing movie reviews with binary sentiment labels (positive/negative). It is widely used for sentiment analysis model benchmarking.
* **Amazon Product Reviews:** This dataset contains reviews of various products sold on Amazon, and the reviews are annotated with sentiment labels.
* **Twitter Sentiment Analysis:** Datasets of tweets labeled with sentiment labels, commonly used for social media sentiment analysis.

Creating a sentiment analysis project involves a combination of programming languages, libraries, and tools. The tech stack includes Python, a popular langauge in NLP for its libraries; NLTK for various NLP tasks; Scikit-Learn for machine learning; TensorFlow or PyTorch for deep learning; Pandas for data manipulation; SQLite or MySQL for efficient data storage; and GitHub or GitLab for version control and collaboration with others.

Here is a list of few NLP projects on sentiment analysis you can start with:

* [Ecommerce product reviews — Pairwise ranking and sentiment analysis](https://bit.ly/47Tyb3t?ref=hackernoon.com)
* [Many-to-One LSTM for Sentiment Analysis and Text Generation](https://bit.ly/3YPoEXf?ref=hackernoon.com)

**Text Classification**



[from freepik.com](https://www.freepik.com/free-icon/chat_14673880.htm#query=text%20classification&position=1&from_view=search&track=country_rows_v2)

Text classification in NLP involves the process of automatically categorizing or labeling pieces of text into predefined categories or classes based on their content and meaning. This task is aimed at teaching computers to understand and organize large amounts of text data, such as emails, articles, or social media posts, by assigning them to specific categories like spam or not spam, topics like sports or technology, and more.

Text classification serves as a cornerstone in information organization by enabling the systematic categorization of textual content. This categorization enables businesses, researchers, and individuals to quickly access, sort, and analyze information.

There are several publicly available datasets that cover a wide range of text classification tasks, such as spam detection, topic classification, and more. Examples include the 20 Newsgroups dataset for topic classification and the Enron email dataset for email categorization.

Creating a text classification project involves assembling a suitable tech stack that leverages the power of NLP libraries and machine learning frameworks. You can use NLTK for various NLP tasks; Scikit-Learn for machine learning; TensorFlow or PyTorch for deep learning; Pandas for data manipulation; SQLite or MySQL for efficient data storage; and GitHub or GitLab for version control and collaboration with others.

Here is a list of few NLP projects on text classification you can start with:

* [Build a Multi ClassText Classification Model using Naive Bayes](https://bit.ly/3YUCIhX?ref=hackernoon.com)
* [PyTorch Project to Build a LSTM Text Classification Model](https://bit.ly/3PeByuw?ref=hackernoon.com)

**Topic Modeling**



[from freepik.com](https://www.freepik.com/free-vector/setup-wizard-concept-illustration_7140756.htm#query=topics%20categorization&position=18&from_view=search&track=country_rows_v2)

Topic modeling in NLP is a technique that involves automatically identifying and extracting the main themes or topics present in a collection of texts. It aims to uncover the underlying structure within the text data by grouping together words that frequently appear together and represent coherent subjects. This helps in gaining insights into the main subjects discussed in the documents and enables various applications like content recommendation, information retrieval, and summarization.

Datasets for topic modeling in NLP include various text sources like news articles, academic papers, social media posts, reviews, blogs, legal documents, and more. These datasets are used to automatically identify and extract main subjects within texts. Depending on the application, datasets may include healthcare records, email archives, or specialized domain-specific data.

To create a topic modeling project, a tech stack may involve Python programming language and libraries like NLTK or spaCy for text processing, Scikit-Learn for machine learning tasks, and Gensim for topic modeling algorithms. Deep learning frameworks like TensorFlow or PyTorch can be used for advanced topic modeling approaches, with Pandas for data manipulation and SQLite or MySQL for efficient data storage. Version control is managed through platforms like GitHub or GitLab, combination of tools that covers the project’s various stages, from data preprocessing to model training and predictions.

Here is a list of few NLP projects on Topic modeling you can start with:

* [Topic modeling using Kmeans clustering to group customer reviews](https://bit.ly/3EdNshZ?ref=hackernoon.com)
* [NLP Project on LDA Topic Modelling Python using RACE Dataset](https://bit.ly/44nyNLW?ref=hackernoon.com)

**Name Entity Recognition**



[From shaip.com](https://www.shaip.com/blog/named-entity-recognition-and-its-types/)

Named Entity Recognition (NER) is an NLP task that involves identifying and classifying specific entities, such as names of people, places, organizations, dates, and more, within text. NER aims to automatically categorize these entities to provide structure and meaning to unstructured text data, enabling information extraction, content analysis, and information retrieval.

NER is used across different fields such as information retrieval,chatbots, financial analysis, healthcare, and news categorization to automatically identify and classify specific entities in text which helps to improve search & content analysis and decision-making in various industries.

Datasets tailored for NER tasks contain text with annotated instances of named entities and their corresponding categories, serving as training and evaluation material for NER models. Common datasets include CoNLL-2003 for English NER, Groningen Meaning Bank (GMB) for English entities, and masakhaNER for african languages.

Creating a Named Entity Recognition (NER) project involves assembling a tech stack that includes Python, NLP tools like spaCy or NLTK for entity recognition, machine learning frameworks such as Scikit-Learn for feature engineering, and deep learning platforms like TensorFlow or PyTorch for neural network-based models. NER-specific libraries like Flair or AllenNLP enhance the process, Together, this stack facilitates the comprehensive development of NER models.

Here is a list of few NLP projects on NER you can start with:

* [NLP Project to Build a Resume Parser in Python using Spacy](https://bit.ly/3Ed0fRI?ref=hackernoon.com)
* [MasakhaNER: Named Entity Recognition for African Languages](https://github.com/masakhane-io/masakhane-ner?ref=hackernoon.com)

**Machine Translation**



[From freepik.com](https://www.freepik.com/free-vector/voice-assistant-translating-into-foreign-languages-voice-activated-digital-assistants-smart-speaker-language-support-internet-things-concept-vector-isolated-illustration_11668667.htm#query=machine%20translation&position=0&from_view=search&track=country_rows_v2)

Machine translation in NLP refers to the automated process of translating text or speech from one language to another using computational techniques and algorithms. This process involves teaching computers to understand the meaning and structure of a source language text(e.g English) and generate an equivalent text in a target language (e. Swahili).

In tourism and travel, machine translation helps people who visit different places and speak different languages. It translates things like menus, signs, and travel guides, making it easier for travelers. In government diplomacy, machine translation helps countries talk to each other by translating important papers and messages. This helps countries work together and understand each other better.

Machine translation datasets contain sentence pairs in different languages to train and test translation models. These datasets include parallel corpora like Europarl and MultiUN user-contributed translations, and more. Custom datasets can be created for specific domains.

The machine translation tech stack includes programming languages like Python, NLP libraries such as spaCy, specialized machine translation frameworks like OpenNMT, pre-trained models like Transformers, data processing tools like Pandas, and alignment/tokenization tools. Deep learning frameworks like TensorFlow or PyTorch are used for training the models. The stack enables the development of translation models, covering data preprocessing, model training, deployment, and evaluation.

Here is a list of few NLP projects on machine translation you can start with:

* [A Machine Translation project that translates text from English to French](https://github.com/3ba2ii/Machine-Translation-NLP?ref=hackernoon.com)
* [English to Italian Neural Machine Translator](https://github.com/balamurugan1603/English-to-Italian-Neural-Machine-Translator?ref=hackernoon.com)

**Question Answering**



[From freepik.com](https://www.freepik.com/free-vector/faqs-concept-illustration_12781055.htm#query=question%20answering&position=15&from_view=search&track=country_rows_v2)

Question Answering (QA) in NLP refers to the automated process of extracting precise answers from a given text or document in response to user-generated questions. QA systems aim to understand the meaning of the questions and the context of the text to locate relevant information and generate accurate answers. These systems can be applied to various domains, such as search engines, customer support, educational platforms, and information retrieval, enabling users to quickly obtain specific information without manually reading through extensive texts.

Datasets used for Question Answering tasks contain pairs of questions and corresponding answers and come in various formats and types to cover different types of questions and texts. Some common types of QA datasets include:

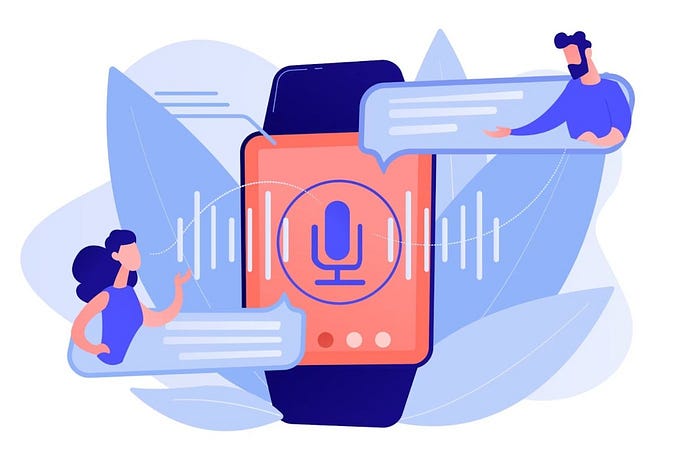
* **SQuAD (Stanford Question Answering Dataset):** A widely used dataset with questions sourced from Wikipedia articles and their corresponding paragraphs containing answers.
* **TriviaQA:** A dataset containing questions from trivia competitions, sourced from Wikipedia, and including evidence documents.
* **NewsQA:** Questions created by humans based on news articles, with corresponding sentences serving as answers.

Creating a QA project involves a specific tech stack that includes using programming languages like Python, and libraries such as spaCy or NLTK for text preprocessing and linguistic analysis. Deep learning frameworks like TensorFlow or PyTorch are utilized for building and training QA models. Specialized QA libraries like Hugging Face’s Transformers provide pre-trained models and tools for QA tasks.

Here is a list of few NLP projects on Question Answering you can start with:

* [Question Answering On SQUAD](https://github.com/ozdemiroorhan/NLP-SQUAD?ref=hackernoon.com)
* [FinBERT-QA-notebooks](https://github.com/yuanbit/FinBERT-QA-notebooks?ref=hackernoon.com)

**Automatic Speech Recognition**



Automatic Speech Recognition (ASR) in NLP refers to the technology that converts spoken language into written text. ASR involves the use of computational algorithms and models to transcribe spoken words from audio recordings or real-time speech into accurate and readable text format. ASR has a wide range of applications, including transcription services (Ref), voice assistants (Amazon Alexa, Apple Siri, and Google Assistant) and more, enabling human-computer interaction through spoken language.

Datasets used for developing Automatic Speech Recognition (ASR) systems consist of paired audio recordings and their corresponding transcriptions in text format. These datasets are crucial for training and evaluating ASR models. Some commonly used ASR datasets include:

* **CommonVoice:** An open-source dataset with multilingual audio recordings and transcriptions contributed by volunteers, used to build ASR models for various languages around the world.
* **LibriSpeech:** This dataset contains audiobooks with aligned transcriptions, providing a diverse range of speech patterns and accents.
* **Custom Created Datasets:** Organizations or communities can create their own datasets by recording speech related to specific domains or industries.

Creating an Automatic Speech Recognition (ASR) project involves a tech stack including programming languages like Python, audio processing libraries such as librosa, specialized ASR libraries like Kaldi or Mozilla DeepSpeech, Nemo, deep learning frameworks like TensorFlow or PyTorch for model development, ASR-specific libraries like SpeechRecognition or vosk for integration, and data augmentation tools like SoX for enhancing the dataset.

Here is a list of few NLP projects on speech recognition you can start with:

* [End-to-End Speech Emotion Recognition Project using ANN](https://bit.ly/3KYKvWC?ref=hackernoon.com)
* [Speech Recognition Macedonian](https://github.com/HristovB/Speech_Recognition_Macedonian?ref=hackernoon.com)

**Conclusion**

NLP project-based learning offers hands-on experience, allowing you to apply theoretical knowledge to real-world situations. This approach promotes critical thinking, problem-solving, and creativity while encouraging collaboration and teamwork. Engaging in projects helps you to gain practical skills in coding, data manipulation, model building, and deployment. Also, it helps to improve your employability and confidence.

It’s important to select projects that resonate with your passions and align with your expertise. Choosing projects that genuinely interest you keeps motivation high and makes the learning experience more enjoyable. Leveraging your existing skills and knowledge ensures a smoother learning curve and a higher chance of success. By aligning projects with your interests and expertise, you’ll not only maximize your learning but also create valuable outcomes that reflect your strengths and dedication.

If you learned something new or enjoyed reading this article, please share it so that others can see it. Until then, see you in the next post!

You can also find me on Twitter [@Davis\_McDavid](https://twitter.com/Davis_McDavid?ref=hackernoon.com).

*One last thing: Read more articles like this in the following links.*

**[How to Web Scrape Using Python, Snscrape & HarperDB](https://python.plainenglish.io/how-to-web-scrape-using-python-snscrape-harperdb-56c53cac876e?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)**

[Store scraped data automatically in the database usin HarperDB](https://python.plainenglish.io/how-to-web-scrape-using-python-snscrape-harperdb-56c53cac876e?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

[python.plainenglish.io](https://python.plainenglish.io/how-to-web-scrape-using-python-snscrape-harperdb-56c53cac876e?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

**[Becoming a Better Data Scientist in 2023](https://python.plainenglish.io/becoming-a-better-data-scientist-in-2023-27928fa14aef?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)**

[5 Tips to Help You Become a Better Version of Yourself](https://python.plainenglish.io/becoming-a-better-data-scientist-in-2023-27928fa14aef?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

[python.plainenglish.io](https://python.plainenglish.io/becoming-a-better-data-scientist-in-2023-27928fa14aef?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

**[Getting a Job as an NLP Engineer: Why and How You Should Find Mentors](https://medium.com/geekculture/getting-a-job-as-an-nlp-engineer-why-and-how-you-should-find-mentors-79772e623228?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)**

[First Important Step in Landing a High Paying Career in Natural Language Processing and AI](https://medium.com/geekculture/getting-a-job-as-an-nlp-engineer-why-and-how-you-should-find-mentors-79772e623228?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

[medium.com](https://medium.com/geekculture/getting-a-job-as-an-nlp-engineer-why-and-how-you-should-find-mentors-79772e623228?source=post_page-----19c8e81c4c58--------------------------------" \t "_blank)

*This article was first published on*[*Hackernoon.*](https://hackernoon.com/7-nlp-project-ideas-to-enhance-your-nlp-skills)

**In Plain English**

*Thank you for being a part of our community! Before you go:*

* *Be sure to****clap****and****follow****the writer! 👏*
* *You can find even more content at*[***PlainEnglish.io***](https://plainenglish.io/)***🚀***
* *Sign up for our*[***free weekly newsletter***](http://newsletter.plainenglish.io/)*. 🗞️*
* *Follow us on*[***Twitter***](https://twitter.com/inPlainEngHQ)***(X***), [***LinkedIn***](https://www.linkedin.com/company/inplainenglish/), [***YouTube***](https://www.youtube.com/channel/UCtipWUghju290NWcn8jhyAw), and [***Discord***](https://discord.gg/XxRS92b2)***.***