PROJECTS


Analytics Vidhya

# NLP

**(Natural Language Processing)**

(Using Python)

# Projects- Natural Language Processing (NLP) Using Python

## Social Media Information Extraction (In-class)

This project is designed to teach you how to extract relevant information such as entities, ngrams, keywords and sentiments from social media data using NLP techniques

## Categorization of Sports Articles (In-class)

In this project, we will learn Document categorization or segregation which is an important NLP task

## SMS Spam Classification (In-class)

In this project, you will learn to preprocess sms text data, feature engineering techniques, and text classification techniques using machine learning and deep learning models.


## Hate Speech Classification

The objective of this task is to detect hate speech in tweets.



Analytics Vidhya

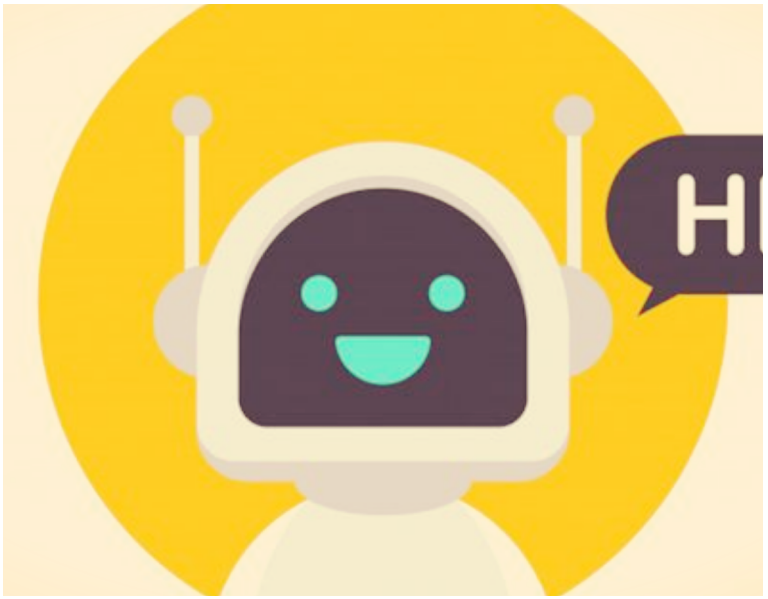
# Projects- Natural Language Processing (NLP) Using Python


## Building Auto Tagging System (In-Class)

In this project, we will build an automatic tagger for the stackoverflow questions.

## Summarization of Customer Reviews (In-Class)

In this project, we will create short summaries of customer reviews on the women's clothing dataset, using sequence-to-sequence modeling.



## Build your first Chatbot (In-Class)

In this project, you will build your first chatbot to search for restaurants online and learn how to use it in a real-world application by deploying it on Slack.

Enroll Now

Analytics Vidhya

# Project- Social Media Information Extraction (In- Class)

# Problem Technique, Statement and Data Description

Enroll Now

- ## Problem Technique:

  Information Extraction

- ## Problem Statement:

  Information Extraction (IE) is a crucial cog in the field of Natural Language Processing (NLP) and linguistics. This project is designed to teach you how to extract relevant information such as entities, ngrams, keywords and sentiments from social media data using NLP techniques. The project highlights the importance of nlp techniques, studied so far, to extract business insights from the text data.

- ## Data Description:

  tweets.csv: ['text'] contains 14940 observations (tweets)

- ## Business Solving Similar Problem:

  Text Data Mining, Knowledge Graph Preparation

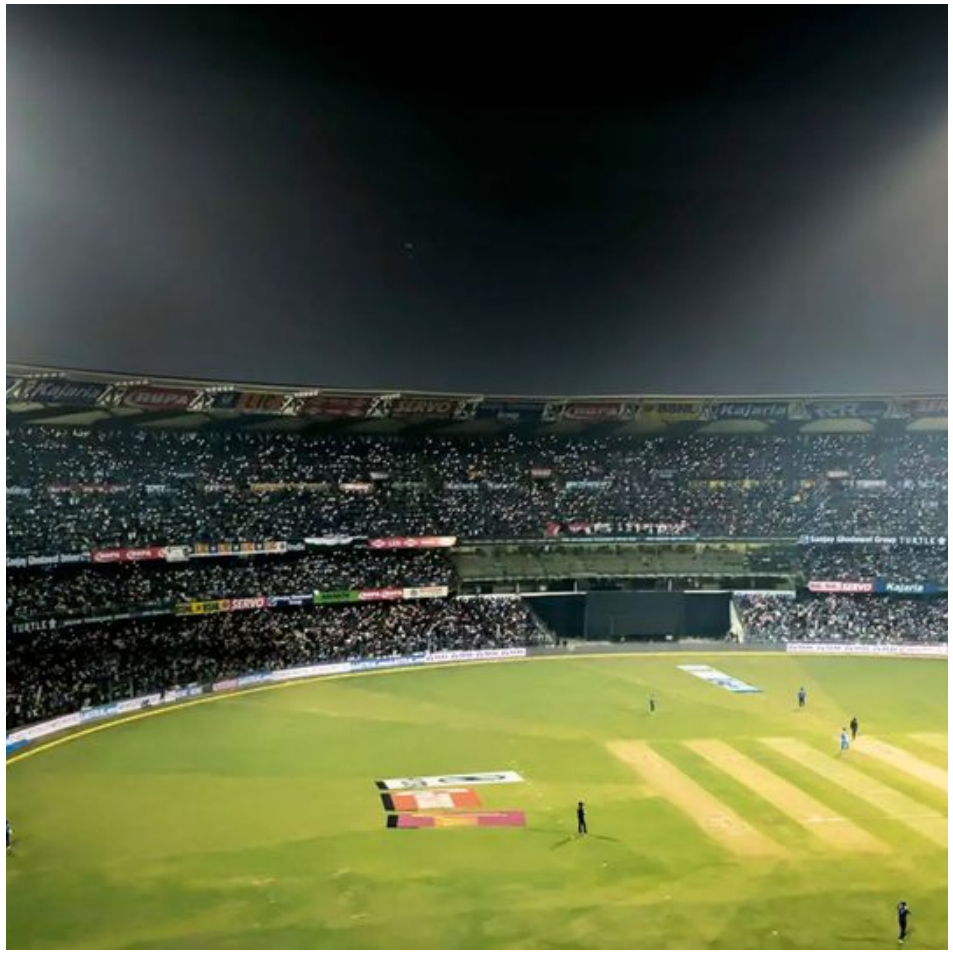


Analytics Vidhya

# Project- Categorization of Sports Articles (In- Class)

# Problem Technique, Statement and Data Description

Enroll Now



- ## Problem Technique:

  Topic Modelling

- ## Problem Statement:

  Document categorization or segregation is an important NLP task which is used across a wide range of industries. In this project, we will learn to segregate sports-news articles using an unsupervised technique called Topic Modelling. We will categorize the articles based on the content of the articles, i.e., similar articles will be grouped together.

- ## Data Description:

  sports.zip: contains 471 text files of sports articles

- ## Business Solving Similar Problems:

  Search Engines, Discovering latent patterns in massive text datasets

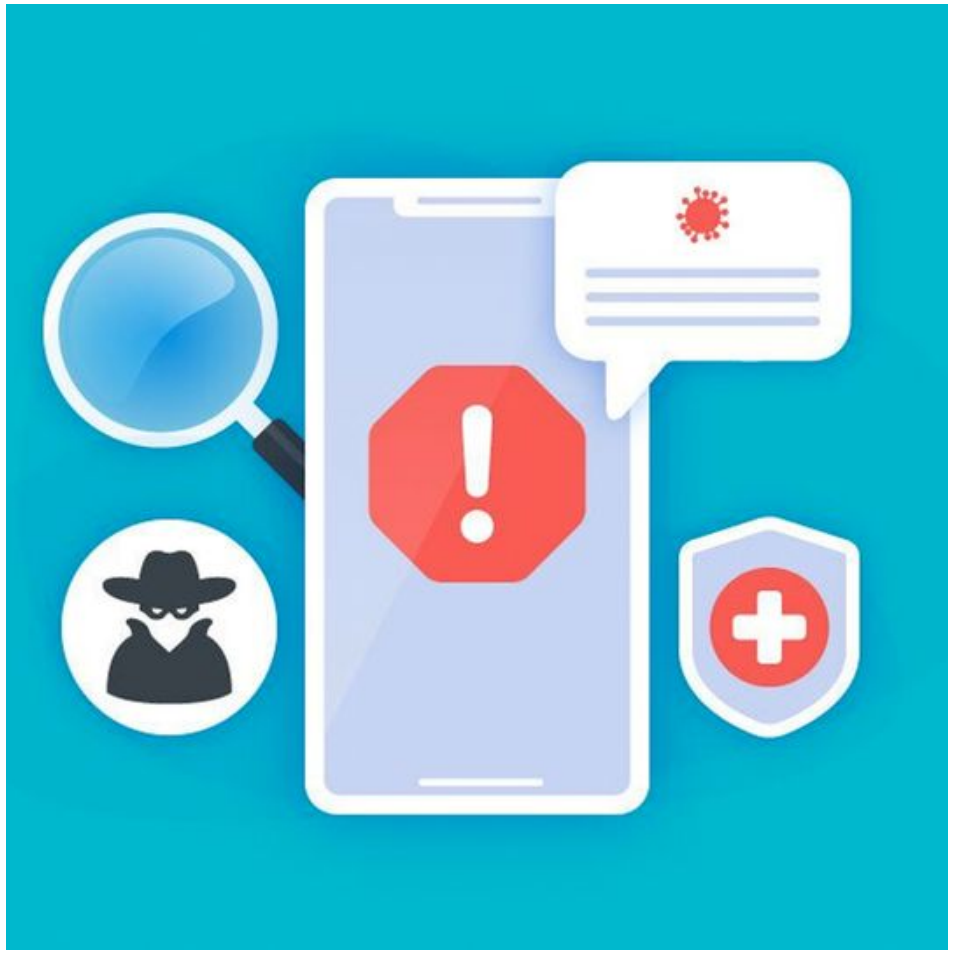


Analytics Vidhya

# Project- SMS Spam Classification (In-Class)

# Problem Technique, Statement and Data Description

Enroll Now



- ## Problem Technique:

  Text Classification

- ## Problem Statement:

  This project is about the classification of SMS text messages as spam or nonspam. In this project, you will learn to preprocess sms text data, feature engineering techniques, and text classification techniques using machine learning and deep learning models.

- ## Data Description:

  spamdata.csv: [‘label’,’text’] contains 5572 observations. ‘text’ contains the sms text and ‘label’ contains the labels “ham” or ”spam”

- ## Business Solving Similar Problems:

  Spam Filtering, Profanity Detection



Analytics Vidhya

# Project- Hate Speech Classification

# Problem Technique, Statement and Data Description

Enroll Now


- ## Problem Technique:

  Text Classification

- ## Problem Statement:

  Hate speech is an unfortunately common occurrence on the Internet. Often social media sites like Facebook and Twitter face the problem of identifying and censoring problematic posts while weighing the right to freedom of speech. The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets.

- ## Data Description:

  train.csv: [‘id’, ‘label’, ‘tweet’] contains 31962 observations

  test.csv: [‘id’, ‘tweet’] contains 17197 observations

- ## Business Solving Similar Problems:

  Spam Filtering, Profanity Detection



Analytics Vidhya

# Project- Building Auto Tagging System (In-Class)

# Problem Technique, Statement and Data Description

Enroll Now


- ## Problem Technique:

  Multi label Text Classification

- ## Problem Statement:

  Automatic tagging of questions on platforms like stackoverflow is quite vital to build a healthy user engagement at the platform. These tags help both, the users seeking solutions to their problems and the experts capable of solving those problems, find the relevant questions easily. In this project, we will build an automatic tagger for the stackoverflow questions.

- ## Data Description:

  auto_tagging_data_v2.h5: [‘Title’, ‘Body’, ‘Tags’] contains 76,000 question-tags pairs

- ## Business Solving Similar Problems:

  Online Content Tagging



Analytics Vidhya

# Project- Summarization of Customer Reviews (In-Class)

# Problem Technique, Statement and Data Description

Enroll Now

- ## Problem Technique:

  Text Summarization

- ## Problem Statement:

  Automatic Text Summarization is a process of generating a concise and meaningful summary of text from multiple text resources such as news articles, blog posts, research papers, customer reviews, emails, and tweets. In this project, we will create short summaries of customer reviews on the women's clothing dataset, using sequence-to-sequence modeling

- ## Data Description:

  Womens Clothing E-Commerce Reviews.csv: ['Title', 'Review Text'] contains 23,486 observations of product-reviews

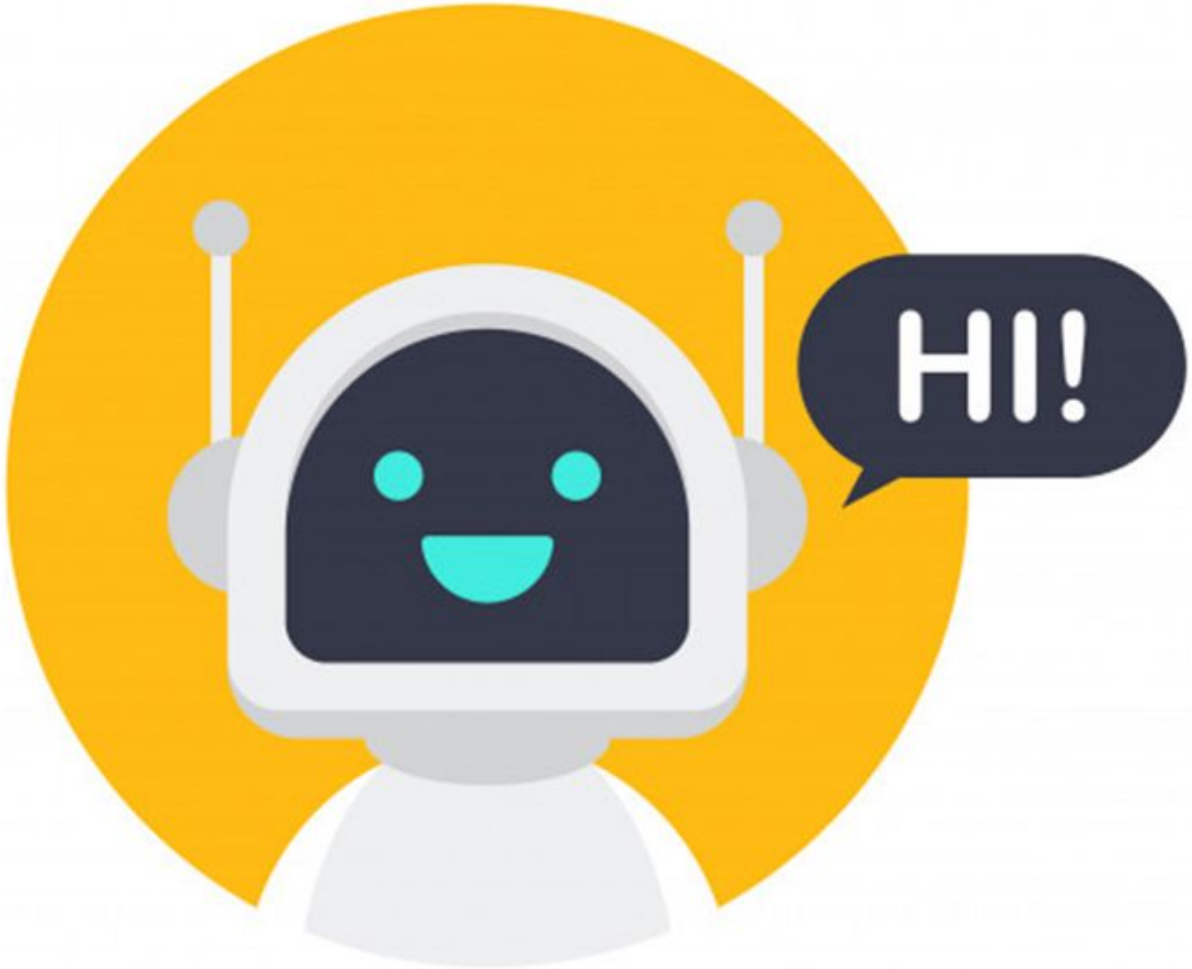
- ## Business Solving Similar Problems:

  Financial Research, Legal contract analysis



Analytics Vidhya

# Project- Build your first Chatbot (In-Class)



# Problem Technique and Statement

Enroll Now


- ## Problem Technique:

  Conversational Agent Development

- ## Problem Statement:

  Chatbots are everywhere today, from booking your flight tickets to ordering food, chances are that you have already interacted with one. In this module, you will build your first chatbot to search for restaurants online and learn how to use it in a real-world application by deploying it on Slack.

- ## Business Solving Similar Problems:

  Customer Support, Lead Generation



Analytics Vidhya


# Thank You

Be the Industry Ready | Enroll Now