**Overview**

Building upon the success of our previous projects—where we analyzed bike rental data, forecasted demand using machine learning, and classified vehicles using CNNs—this next phase focuses on analyzing customer review data. The objective is to analyze customer reviews, extract insights, and understand sentiment to improve BikeEase’s services.

BikeEase receives thousands of customer reviews and feedback across multiple platforms. However, manually analyzing this data is inefficient. The goal of this capstone is to develop an NLP-powered sentiment analysis system that automatically classifies reviews as positive, neutral, or negative. Additionally, we will perform topic modeling to uncover key themes in customer feedback.

**Project Statement**

Develop an end-to-end NLP pipeline to analyze customer reviews for sentiment classification and key topic extraction. This system will help BikeEase identify customer pain points and areas of improvement.

**Input dataset:** [Dataset](https://drive.google.com/drive/folders/13-g3jxhPR0btN_s77KbcempcFXZ0RoqT?usp=drive_link)

[Links to an external site.](https://drive.google.com/drive/folders/13-g3jxhPR0btN_s77KbcempcFXZ0RoqT?usp=drive_link)

**Steps to Perform**

**Task 1: Data Collection & preprocessing**

* + Collect and clean customer reviews from a given dataset (or scrape data if available)
  + Perform text cleaning (lowercasing, removing punctuation, stopword removal, lemmatization)
  + Tokenize and vectorize the text if required

**Task 2: Sentiment analysis**

* + Build a **sentiment classification model** (positive, neutral, negative) using:
    - Traditional models: Logistic Regression, Naïve Bayes
    - Deep learning models: LSTMs, Transformers (BERT)
  + Evaluate models using accuracy, F1-score, and confusion matrix