

# ISYE 6740 - Spring 2024

## Ryanair Passenger Experience Review

### Term Project 114

Ashish Puri  
apuri61@gatech.edu

Jagannath Banerjee  
jbanerjee7@gatech.edu

Piyush Shivrain  
pshivrain3@gatech.edu

## 1 Project Description / Problem Statement

In this project, we aim to conduct a comprehensive analysis of passenger experience reviews for Ryanair, spanning from 2012 to 2024. Ryanair, being one of Europe's leading low-cost airlines, garners a vast amount of feedback from its passengers, covering various facets of their travel experience. Through this analysis, we intend to extract insights that can benefit both the airline and potential travelers.

## 2 Data

Dataset has been taken from Kaggle <https://www.kaggle.com/datasets/cristaliss/ryanair-reviews-ratings?resource=download>. Dataset contains the following information:

- Passenger-reported ratings on seat comfort, cabin crew service, food & beverages, ground service, and overall value for money.
- Detailed insights into the types of travelers, such as leisure, business, or family.
- Information on aircraft types, seat types, routes flown, and dates of travel.
- A breakdown of passenger nationalities and trip verification status.

## 3 Research Questions

This project aims to address several key research questions to gain a deeper understanding of passenger experience with Ryanair using diverse array of opinions and ratings provided directly by passengers.

- What are the most common words and sentiments expressed in the reviews?
- How do passenger ratings distribute across different aspects of the travel experience?

- Are there any discernible trends or patterns over time in passenger feedback?
- Can we identify specific areas for improvement based on passenger sentiments?
- What are the prevalent themes/topics discussed in the reviews, and how do they correlate with sentiments?

## 4 Methodology

To analyze the data, we will employ a combination of descriptive statistics, exploratory data analysis (EDA), and advanced data mining techniques.

### EDA:

- Conduct word frequency analysis to identify common words in reviews.
- Perform sentiment analysis to gauge overall sentiment (positive, negative, neutral).
- Utilize word clouds for visualizing word frequencies.
- Analyze the distribution of star ratings given by customers.

### Modelling:

- **Topic Modeling:** Identifying topics/key areas of improvement present in the reviews using techniques like Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF).
- **Sentiment Classification:** Build sentiment classifier model to predict sentiment labels (positive, negative, neutral) based on reviews.

To construct the above mentioned topic modelling and sentiment analysis , we would first do text preprocessing and then feature extraction as described below:

- **Text Preprocessing:**

Tokenization: Break down the text into individual words or phrases (tokens).

Lowercasing: Convert all text to lowercase to ensure consistency.

Removing Stopwords: Eliminate common words (e.g., 'and', 'the') that don't carry significant meaning.

Stemming or Lemmatization: Reduce words to their root form to normalize the text (e.g., 'running' to 'run').

Handling Contractions and Abbreviations: Expand contractions ('won't' to 'will not') and handle abbreviations for consistency.

Removing Punctuation and Special Characters: Eliminate non-alphanumeric characters that don't contribute to the analysis.

Handling Numerical Data: Decide whether to treat numbers as text or convert them to a standard format.

- **Feature Extraction:**

Bag-of-Words (BoW): Represent each document as a vector where each dimension corresponds to a word and its value represents the word's frequency.

Term Frequency-Inverse Document Frequency (TF-IDF): Weigh the importance of words in a document relative to their frequency across all documents.

Word Embeddings: Convert words into dense vectors to capture semantic relationships (e.g., Word2Vec, GloVe).

N-grams: Consider sequences of adjacent words (bi-grams, tri-grams) to capture context and phrases.

Topic Modeling Features: Extract topics from text using techniques like Latent Dirichlet Allocation (LDA) and use the topic distribution as features.

Named Entity Recognition (NER): Identify and extract named entities such as locations, organizations, and people.

Syntax-based Features: Extract syntactic features like parts-of-speech tags or syntactic dependencies.

In addition, if we see fit, we may include the following techniques to optimize the final output

- **Additional Methodology:**

Cross-Validation: Utilize techniques like k-fold cross-validation to assess model performance robustly.

Hyperparameter Tuning: Fine-tune parameters of models using techniques like grid search or random search.

Model Ensemble: Combine predictions from multiple models to improve performance.

Error Analysis: Investigate misclassified instances to understand model limitations and potential areas for improvement.

Interpretability: Use techniques like SHAP (SHapley Additive exPlanations) to interpret model decisions and extract actionable insights.

Dynamic Topic Modeling: Explore methods to track evolving topics over time, considering the changing nature of passenger feedback.

## 5 Evaluation

We can evaluate the performance of advanced data mining techniques such as topic modeling and sentiment classification in identifying prevalent themes and sentiments within the reviews, considering their applicability to real-world scenarios and potential benefits for Ryanair and its passengers. We will incorporate accuracy metrics and confusion matrix analysis to assess the performance of the sentiment classifier model. By calculating accuracy, precision, recall, and F1-score, we can quantify the model's ability to correctly classify sentiments within the reviews. Additionally, we can gauge the overall coherence and actionable insights derived from the project findings, assessing their potential impact on enhancing the passenger experience with Ryanair.

## References

- [1] "Ryanair Passenger Experience Reviews." Wwww.kaggle.com, [www.kaggle.com/datasets/cristaliss/ryanair-reviews-ratings?resource=download](https://www.kaggle.com/datasets/cristaliss/ryanair-reviews-ratings?resource=download). Accessed 11 Mar. 2024
- [2] TNMT. (n.d.). Passenger Frustration with Airlines. Retrieved from <https://tnmt.com/passenger-frustration-with-airlines/>
- [3] Dike, S. E., Davis, Z., Abrahams, A., Anjomshoae, A., Ractham, P. (2023, March 21). Evaluation of passengers' expectations and satisfaction in the airline industry: An Empirical Performance Analysis of online reviews. Benchmarking: An International Journal. <https://www.emerald.com/insight/content/doi/10.1108/BIJ-09-2021-0563/full/html>
- [4] Ali, A. (2023, July 13). Airline reviews sentiment analysis by NLP and POWERBI. LinkedIn. <https://www.linkedin.com/pulse/airline-reviews-sentiment-analysis-nlp-powerbi-ahmed-ali/>