Assessing Social Media Communications Of Airlines

Author: Mohammad Fayazi
Supervisor: Jerry Spanakis
Email: m.fayazi@student.maastrichtuniversity.nl

***Abstract*—This study aims to understand how airlines' social media posting frequency and content impact customer engagement rates. The investigation compared different airlines' posting frequencies and changes during two time periods, normal times and COVID-19 lockdown restrictions, and their consequent impact on brand engagement and popularity. The findings indicated that an increase in posting frequency corresponded to higher brand engagement for both large and small airlines. However, the effect on brand popularity was less straightforward, particularly for smaller airlines.**

In addition, the study investigated the relationship between specific features of airlines' social media posts, such as sentiment, topics, length, timing, and type of posts, and the level of engagement they received. By applying machine learning algorithms (Linear Regression, Support Vector Regression, and Neural Networks) to predict Post Engagement Rates, it was found that the topic of posts was a significant driver of engagement. Moreover, a straightforward model like Linear Regression outperformed more complex models, suggesting simplicity could be an advantage in predicting engagement rates.

Lastly, text vectorization techniques, including BERT, FastText, and TF-IDF, were compared for their effectiveness in grasping the semantic content of airline posts and predicting their engagement rates. BERT, combined with Support Vector Regression, provided the best performance among these techniques. However, the results were comparable to those obtained from Linear Regression using the "Topic" feature, suggesting the relative effectiveness of both sophisticated and simple techniques. The study provides airlines with valuable insights into optimizing their social media strategies for maximum customer engagement.

***Index Terms*—Airline Communication Strategies, Posting Frequencies, Facebook Content Feature Analysis, Social Media Analytics, Facebook Post Engagement, Airline Industry,Text Vectorization, Sentiment Analysis, Linear Regression, Support Vector Regression, Neural Networks, TF-IDF, FastText, BERT, Time-Series Analysis, Content Features, Customer Engagement**

I. Introduction

In an increasingly digitized world, social media platforms have risen to play an important role in the daily operations of many businesses [1]. Airlines, like other service-oriented industries, have leveraged social media as a primary means of communication with their customers, providing information, promotions, and even handling customer service inquiries [2]. Therefore, understanding social media dynamics and their impact on customer's engagements becomes a crucial aspect of any strategic decisions in the airline industry. This thesis aims to explore these dynamics, particularly focusing on airlines' Facebook communications.

The problem statement that this thesis addresses is as follows: In the context of the airline industry's reliance on Facebook as a primary mode of communication, there is a need for a robust understanding of how airlines' posting strategies, specific for different features influence the engagement rates of their posts. This complex issue has significant implications on the effectiveness of airlines' social media strategies, as it directly impacts customer engagement, brand popularity, and ultimately the success of their digital marketing efforts [3].

The first research question this thesis addresses concerns "How have the posting frequencies of different airlines changed over time? & what impact has this had on their social media engagement and popularity?". The aim is to analyze trends in posting frequency and quantify the effects on various engagement metrics such as likes, comments, shares, follower count, and total interactions. Spanning from the beginning of 2020 to the initial months of 2023, the selected time frame provides an opportunity to investigate this issue under varying circumstances, including periods of standard operation and times significantly impacted by reduced air travel due to COVID-19 lockdowns. This differentiation between periods lead us to presume on the potential effect of posting strategies on airlines' engagement and popularity under diverse conditions.

The second research question focuses on how specific features of airlines' posts are correlated with the level of engagement they receive. Various aspects of Facebook posts, including sentiment they convey, post length, timing, type, content and the topics they discuss, are dissected. For each feature, its relationship with the engagement metric, was quantified and defined by the combination of likes, comments, and shares a post receives. Then, machine learning techniques, including Linear regression, Support Vector Machines (SVM) and Neural networks, was applied to predict the engagement rate of future posts based on these features. This research provides an exploratory analysis of which features (or combination of features) have the most substantial influence on the engagement rate of posts.

In the final segment, our third research question seeks to examine the capability of advanced text vectorization methods, specifically BERT, FastText, and Tfidf, in understanding the semantic content of airline posts and predicting their engagement rates. These techniques enables the transformation of text data into numerical vectors that can be effectively processed by machine learning algorithms. By comparing the performance of these models against the features identified in the second part of our study, we aim to understand if these sophisticated NLP methodologies can better capture the nuances of the text and provide more accurate predictions of post engagement. Ultimately, this exploration serves to expand the breadth of tools available for the quantitative analysis of social media engagement, potentially offering more refined strategies for maximizing the impact of airlines' communication efforts.

Overall, the objective of this thesis is to provide a comprehensive analysis of airlines' social media strategies on Facebook and their impact on customer engagement. We aspire to provide useful insights that could inform and improve future communication strategies of airlines.

II. Literature Review

To further our understanding of the critical dynamics of the airlines' social media strategies and customer engagement, a review of the existing literature provides insights into current researches and advancements in this domain. The focus is on studies examining the role of content components, interactivity cues, media richness, sentiment analysis, and customer engagement behaviors in social media interactions. These works will highlight our exploration and contextualize our findings in the broader academic discourse.

An empirical study by Moran, G., Muzellec, L. and Johnson, D. (2020, [4]) provides a robust examination of consumer-brand engagement on Facebook. The researchers investigate how content components, interactivity cues (calls to action), and media richness (e.g., video, photo, and text) influence users' behavioral responses, such as clicks, likes, shares, and comments. Analyzing 757 Facebook-based brand posts from a media and entertainment brand, they uncover a positive relationship between both interactivity cues and media richness content components in increasing consumer-brand engagement outcomes. The study further reveals that interactivity cues enhance all forms of engagement behavior, with visual imagery (photos and videos) attracting the most consumer responses.

The effectiveness of social media communication strategies is examined by Prados-Peña et al. (2022, [5]) using two Spanish airlines' Facebook communications during the COVID-19 pandemic. Their findings highlight the relationship between content types and variables such as brand popularity, customer brand engagement, and virality. This reinforces the importance of strategizing content delivery to stimulate positive engagement and reactions.

Regarding the customer engagement behaviors (CEB) on social media, Ajiboye, Harvey, and Resnick (2019) present a comprehensive review of the antecedents of CEB. Key influencers of engagement include social links, ownership-value, information seeking, involvement, and functionality. These findings illustrate the multidimensional nature of user engagement on social media, illuminating how businesses can potentially enhance interaction on these platforms (CEB, 2019 [6]).

In an exploration of Chinese destination management organizations (DMOs), [Zhenxing Mao,Dongjie Li,Yang Yang,Xiaoxiao Fu &Wan Yang, 2019, [7]] studied how these entities use social media to attract the interest of international travelers, focusing particularly on the role of post-related factors in driving user engagement. After reviewing over 6000 Facebook posts over a year, they found that interactive content and post length significantly positively affected user engagement. Notably, post length was found to be particularly influential when combined with interactive content, implying that interactivity and length can interact to further boost engagement.

In terms of linguistic shifts, Kulkarni et al. (2015, [8]) propose an innovative approach for tracking and detecting statistically significant changes in word usage and meaning over time. This method reveals compelling patterns of language change over different timeframes and mediums, suggesting a strong potential for its application in tracking sentiment shifts.

Dolan et al. (2015, [9]) further investigates into social media engagement behaviors, offering a uses and gratifications perspective. They propose a theoretical model outlining how an organization can encourage positively valenced engagement behavior through social media content and reduce negatively valenced behavior. Their typology of social media engagement behavior serves as a guide for examining relationships between social media content and engagement.

The challenge of real-time sentiment analysis in machine learning due to data scarcity and sudden sentiment changes is explored by the ACM international conference on Web search and data mining. They propose methods to generate labels and manage concept drift using findings from social psychology. The focus of the paper is on terms that spike in social streams instead of static text representation, which led to the detection of new informative features capturing sudden changes on sentiment stream caused by real-world events. These strategies contributed to high accuracy in live reactions analysis despite the absence of human-generated supervisory labels (WSDM, 2014 [10]).

The creation of effective SVM classifiers for detecting message and term-level sentiment is discussed by authors of an ACL Anthology paper. Their classifiers performed excellently, demonstrating the value of features like surface-form, semantic, and sentiment, as well as the usefulness of large word-sentiment association lexicons in improving the accuracy of sentiment detection (S13-2053, 2013 [11]).

Overall, the literature underscores the promise and complexity of sentiment analysis, customer engagement, and data mining in social media. These works collectively provide a solid foundation for this thesis, which seeks to further contribute to the body of knowledge in these areas.

III. Data set Analysis:

This section will consider the dataset that was collected and an overview of the information that it offers:

A. Data Collection

The dataset used in this study was obtained from CrowdTangle, a public insights tool owned and operated by Facebook, designed to provide access to social media data for research purposes [12]. CrowdTangle provides access to a vast amount of information related to Facebook pages and their posts, offering a broad range of metrics such as the number of likes, comments, shares, and other interaction data.

B. Data Description

This research focused on European airlines, comprising 90 different airlines from 40 countries, resulting in nearly 33,000 data points collected over the time period from the beginning of 2020 to the beginning of 2023. The data collected provides insights into the detailed information for each post, such as user name, likes at posting, followers at posting, post creation time and date, post type, total interactions, likes, comments, shares. Additionally, it includes information related to the post content, such as messages, image text and description; forming a rich dataset to explore the relationship between posting frequencies, post characteristics, and user engagement.

The timeline of our dataset spans from the start of 2020 through to the beginning of 2023. In terms of post type, the dataset predominantly comprises photo posts, with other post types including videos, text, and links (figure1).

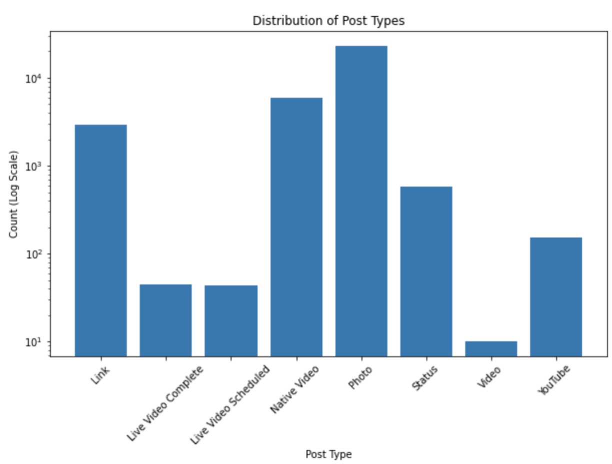

Fig. 1. Post Types

The frequency of posts by airlines within the dataset displayed a significant level of fluctuation over the given time period. A key observation made was the significant drop in posting frequency during the period from March 2020 through September 2020. This drop is in accordance with the global spread of the COVID-19 pandemic and the lockdown restrictions that were imposed worldwide, impacting the regular operations of these airlines.

Additionally, another noticeable drop in posting frequency is observed after August 2022. However, it's important to note that this decrease may be caused by the incomplete data for the latter part of the year, rather than an actual decrease in posting activity (figure2).

It is worth noting that, in the context of our study, we only considered posts that were in English, which accounted for roughly half of the posts in our dataset (approximately 17,000

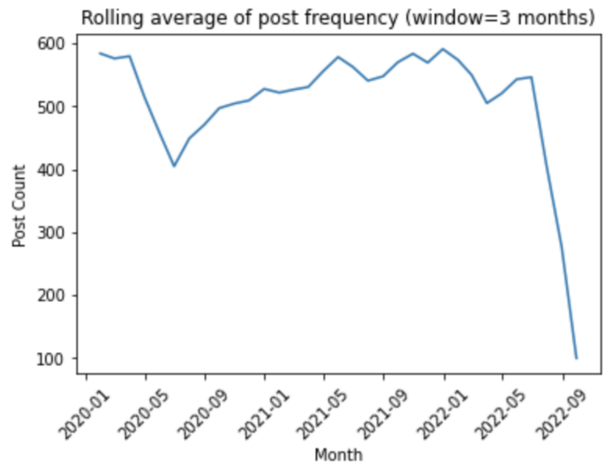

Fig. 2. Post Frequency over time

posts). These posts form the basis of the text and sentiment analyses performed in this study.

IV. Methods

This section is divided into three subsections, each corresponding to the methodology employed to address a distinct research question. The initial subsection presents a comprehensive examination of the statistical techniques employed to resolve the first research question. With regards to the second research question's focus on the characteristics and details of each post's text Thus, the second subsection will incorporate a contextual evaluation of individual posts. Lastly, the third part is dedicated to the vectorisation methods that were used to capture the semantics of the posts.

A. Statistical Methods: Analysing Posting Frequencies

1) Statistical Analysis: Initiating the statistical analysis of the dataset, posts were grouped according to their corresponding airlines, followed by a counting the number of posts for each airline during the study period. This count was arranged in a descending order and airlines contributing fewer than 36 posts were omitted, which equates to those with less than one post per month over the three-year duration. This procedure aimed to make the analysis more robust, despite leading to a reduction in the original size of the dataset.

Next, the average number of followers at the time of posting for each airline was calculated. Airlines with follower counts in the top 20 percentile were considered 'bigger airlines' and those below, 'smaller airlines'. The dataset was further adjusted by categorizing airlines based on their size, thus enabling differentiated analyses for bigger and smaller airlines(figure3).

Two distinct periods were considered for the analysis - period 1 from March 2020 to September 2020 (during which the lockdown restrictions were imposed), and period 2 from September 2020 to June 2021 (during the normal times, when traveling was going back to the normal conditions). Posting

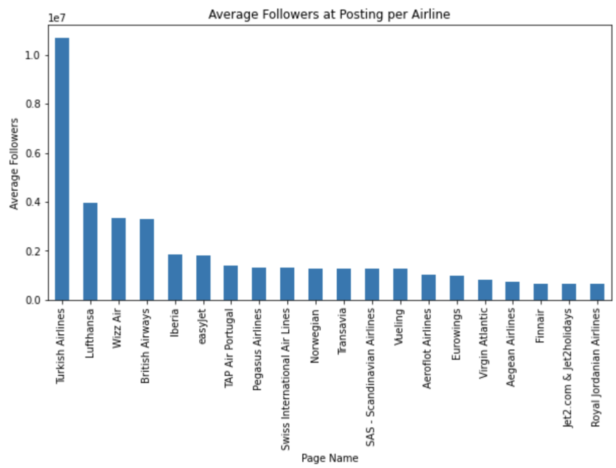

Fig. 3. Airline's Followers

frequencies for both these periods were calculated for each airline.

Independent t-tests were conducted for both the bigger and smaller airlines to compare their mean posting frequencies between the two periods. A significance level of 0.05 was used as the threshold. Results were sorted by p-value to identify the airlines with the most significant changes in posting frequency.

The same periods were then analyzed using the Wilcoxon rank-sum test, a non-parametric test that does not assume normal distribution [13]. This test, which takes a different approach to identify if there were statistically significant changes in the posting frequencies between the two periods, was conducted to ensure us from the result of t-test.

In both the t-tests and the rank-sum tests, airlines for which there were significant changes in posting frequencies between the two periods were found (figure4).

It's important to note that airlines with mean posting frequencies less than 10 in either of the periods were excluded from the tests, and that airlines needed to have made at least 10 posts in both periods to be included in the rank-sum tests.

Additionally, graphical representations of the posting frequencies of the airlines identified by the rank-sum test have been provided in the Appendix of this study (figures 22 23 24 25 26 27 28 29 30). This additional visual aid assists in comprehending the scale of changes in posting frequencies for these airlines.

After analysing the airlines based in their size (bigger airlines and smaller ones) and whether their posting frequencies had changed significantly during the two time periods, 6 different categories of airlines were obtained to be evaluated, namely:

- all airlines with significantly different posting frequencies during the time periods (we call them "overall significant change")
- all other airlines which did not change their posting frequencies ("overall non-significant change")
- big airlines with significantly different posting frequencies during the time periods ("big significant change")

```
Significant changes for bigger airlines:
```

Page Name	W Statistic	p-value
Wizz Air	-2.43975	0.014697
British Airways	-2.14698	0.031795

```
Significant changes for smaller airlines:
```

Page Name	W Statistic	p-value
Icelandair	-2.000595	0.045436
Eurowings	2.146980	0.031795
Vueling	-2.732520	0.006285
Royal Jordanian Airlines	-2.634930	0.008415
Jet2.com & Jet2holidays	-2.732520	0.006285
Austrian Airlines	-2.244570	0.024796
Czech Airlines	2.000595	0.045436
Atlantic Airways	-2.244570	0.024796
Azores Airlines	-2.390955	0.016805

Fig. 4. Arilines with significant change in posting frequency

- big airlines which did not change their posting frequencies ("big non-significant change")
- small airlines with significantly different posting frequencies during the time periods ("small significant change")
- small airlines which did not change their posting frequencies ("small non-significant change")

The performance of each of the above mentioned categories will be evaluated according to the evaluation methods below:

2) Evaluation Methods: Brand Popularity & Engagement: The brand popularity and engagement scores are calculated based on the metrics provided in the dataset. Prados-Pena et al. (2022 [5]) explains how, brand engagement, popularity and virality can be measured, using the number of likes, comments and shares that they received respectively. Nevertheless, our data set offers a more comprehensive information of the brands and their posts, including factors like the number of followers and interactions. Therefore, an expanded version of the evaluation approach presented in Prados-Pena et al (2022) was employed. The brand popularity is a weighted sum of total interactions, followers at posting, and likes at posting. A score is calculated for each post, and the average score is reported for a given time period.

To control for different scales of the contributing features, we scale the popularity score by the maximum possible score that can be achieved. This provides us with a relative popularity measure, which is useful for comparing across different airlines or time periods.

$$\text{Brand Popularity Score}_i = \frac{w_1 \cdot \text{Interactions}_i + w_2 \cdot \text{Followers at Posting}_i}{w_1 \cdot \text{Max Interactions} + w_2 \cdot \text{Max Followers at Posting}} + \frac{w_3 \cdot \text{Likes at Posting}_i}{w_3 \cdot \text{Max Likes at Posting}} \quad (1)$$

Brand engagement is calculated in a similar manner, with weights assigned to likes, comments, and shares of a post. The total engagement for a post is divided by the number of followers at the time of posting to obtain the engagement rate. Again, this score is averaged over the given time period. To account for varying scales, the MinMaxScaler from the sklearn library, which scales the data to a range between 0 and 1 was used [14].

$$\text{Engagement Score}_i = \frac{v_1 \cdot \text{Likes}_i + v_2 \cdot \text{Comments}_i + v_3 \cdot \text{Shares}_i}{\text{Followers at Posting}_i} \quad (2)$$

B. Contextual Methods: Characteristics of Posts

In this section, the methodology to quantitatively analyze the relationship between specific features of airlines' posts and the level of engagement they receive, will be analysed. The methods used involve various data processing, feature extraction, and machine learning techniques. It begins with an extensive text pre-processing phase, and then extracting features from the processed text using Sentiment Analysis, Post Length calculation, Time Series Analysis, and analysis of Post Type. Furthermore, Topic Modelling is conducted to understand the themes present in the posts. To predict the engagement rate, several machine learning regression models were employed, namely Linear Regression, Support Vector Regression, and a Deep Learning model. Finally, the Post Engagement Rate is used as an evaluation metrics to measure the efficacy of our models and the impact of various features on engagement.

1) Text Processing: In order to analyse the textual characteristics of each posts, the messages of each airline's post were collected. These messages contain the text for all posts types (including videos or images and etc). An essential step in preparing this data for analysis is the text processing phase, which involved several stages, namely:

1) HTML tag removal: To clean the text data, HTML tags present in the text were removed using a regular expression (regex) pattern that matches HTML tags. This process is essential to eliminate any potential noise in the data stemming from these tags.
2) URL removal: As with HTML tags, URLs in the text were removed to avoid any noise they might introduce in the analysis. A regex pattern was used to identify and remove URLs from the text data.
3) Emoji conversion: Emojis were converted into textual descriptions using the emoji Python package. This process translates an emoji into a string description of the emoji. For example, the heart emoji becomes 'red_heart'.
4) Contraction expansion: Contractions (e.g., "it's" to "it is") were expanded to their full form regex patterns (e.g., "can't" to "cannot"). This step aids in the normalization of the text.
5) Custom text pre-processing: A custom text pre-processing function was applied to the text data. This function performed several operations:
   - The text was tokenized, and each token was converted to lowercase and lemmatized using the spaCy English model [15].
   - Stop words and punctuation were removed from the tokens. Stop words are commonly used words that do not carry significant meaning and are often removed in text processing to reduce noise.

These pre-processing steps are crucial in preparing the text data for the subsequent stages of analysis, enhancing the accuracy and effectiveness of the methods applied. They helped to reduce noise, normalize the data, and extract meaningful textual features.

2) Features Extraction:: Upon completing the text processing, the first feature derived from the post messages is the sentiment being communicated to the audience. Whether the posts are formulated with a positive, neutral or negative senses, can have significant impact on their audience and consequently on the engagement rate that they cause.

Sentiment analysis was performed using two different techniques: DistilBERT model, a Transformer-based model, and VADER sentiment analysis, a lexicon-based sentiment analysis method.

- DistilBERT Sentiment Analysis: DistilBERT (a lighter version of BERT) is a transformer-based Machine Learning model, pre-trained on a large corpus of English text. It is fine-tuned for sequence classification tasks. First, the tokenizer from the 'distilbert-base-uncased' model was used to encode the processed text messages. The encoded text was restricted to a maximum length of 512 tokens with longer text being truncated. The encoded inputs were then processed by the DistilBERT model to predict the sentiment [16].
- VADER Sentiment Analysis: In addition to the DistilBERT sentiment analysis, VADER (Valence Aware Dictionary and sEntiment Reasoner) was employed. VADER is a lexicon and rule-based sentiment analysis tool. It is specifically attuned to sentiments expressed in social media and calculates the sentiment of text data based on a predefined dictionary of positive and negative sentiments. Using the VADER sentiment analysis tool, polarity scores were calculated for each processed text message [17].

After implementing both techniques, the analysis found that positive sentiments were present in 12,503 instances, neutral sentiments in 1,967 instances, and negative sentiments in 841 instances.

In order to include the sentiment information in machine learning models, these categorical variables need to be converted into numerical format, as machine learning algorithms work better with numerical data. To do so, a pre-processing technique where the string sentiment labels are transformed into a number was employed, making the data suitable for use in an algorithm while retaining the important categorical information.

The next feature explored was the length of the posts. The length of a post can play a crucial role in its engagement rate, since it can impact readability and user experience. mao (2020, [7]) proved that the length and type of the posts has a direct relation with the engagement factors of audiences. In this paper the length of each processed post was computed and stored as a feature for prediction models. It is worth to mention that the length of the posts are based on the number words they are composed of, instead of number of characters. The Post Length distribution can be observed below 5:

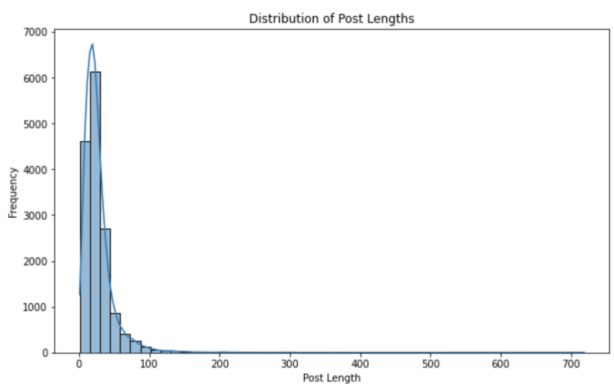

Fig. 5. Post Length Distribution

Additionally, the study conducted a comprehensive time-series analysis on the engagement rates of airlines' social media posts, which included trend, hourly, and monthly analysis. This examination revealed patterns and fluctuations in engagement rates over time.

A rolling mean (moving average) with a window size of seven days was used to reduce the noise and highlight the underlying trend in the post engagement rates. This method, complemented by the removal of outliers using the Interquartile Range (IQR) method, helped clarify the short-term fluctuations and provide a more accurate picture.

The trend analysis calculated and visualized the average engagement rate over time for all major airlines. The use of a regression plot gave a clearer view of how engagement rates change over time. (figure6).

The hourly analysis extracted the post creation time and calculated the average engagement rate for each hour of the day. This offered insights into the optimal posting times for increased engagement. (figure7).

Finally, the monthly analysis converted post creation dates into months and calculated the average engagement rate for

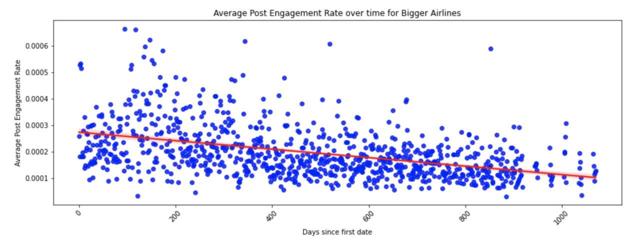

Fig. 6. Big airlines Engagement Change

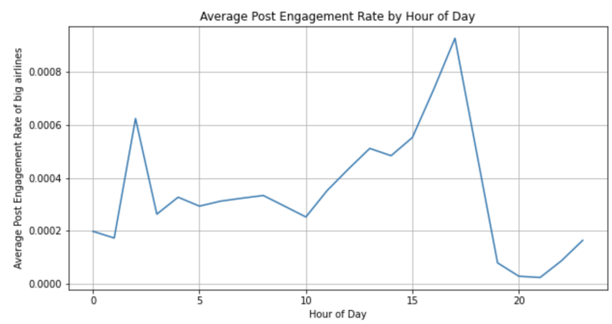

Fig. 7. Engagement rate based on hour of publishing

each month. This analysis revealed how engagement rates might fluctuate due to factors like seasonality across different months. This temporal analysis overall helps airlines strategize their posts timing for maximum engagement. (figure8)

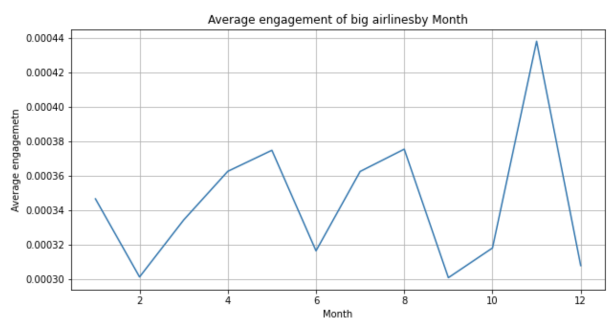

Fig. 8. Engagement rate based on the month of publishing

Further, the effect of post types on engagement rates was also investigated. The analysis revealed which types of posts tend to generate higher engagement rates, providing valuable insights for airlines to enhance their engagement strategy. Similar to the sentiment analysis, these categorical post types were transformed into a numerical format to include them in the predictive machine learning models.

Through the below visualization (figure9), it becomes easier to discern which types of posts tend to generate higher engagement rates. This insight can guide airlines to focus more on creating the types of posts that are likely to engage more audience.

Finally, the last feature examined was the topic of the posts. The study divided the dataset into three subsets based on sentiment labels and performed topic modeling to discover

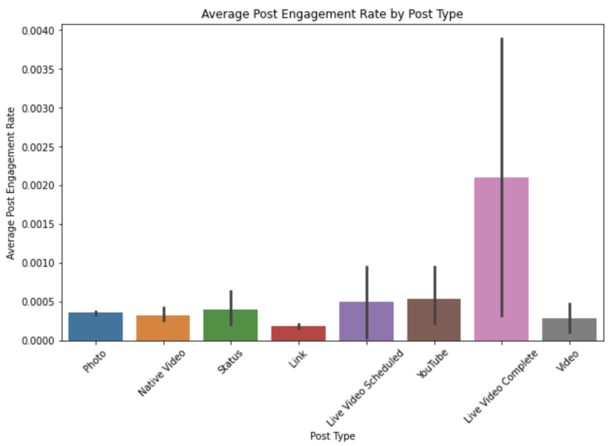

Fig. 9. Engagement rate based type of posts

themes within the posts. By taking sentiment into account, this procedure contributed to a comprehensive understanding of the themes present in the airlines' Facebook posts and how they are expressed across different sentiment categories.

The topic models were created for each sentiment subset using the BERTopic library. BERTopic is a tool for topic modeling with BERT embeddings, which leverages the semantic meaning of words and sentences to improve the understanding of the topics [18]. Each of the topic models was fit to the respective subset of processed text. Finally, these categorical topics were encoded and concatenated back to the original dataframe.

3) Prediction Models: With all these features extracted and prepared, the next phase involves predicting the engagement rate of airline Facebook posts. This phase included the implementation of several predictive models including Linear Regression (a basic and widely-used statistical technique that predicts an outcome variable (y) based on one or more predictor variables (X)), Support Vector Regression (SVR) (a type of Support Vector Machine (SVM) that is used for regression tasks. The major advantage of SVR over other techniques is its ability to handle non-linear relationships using different types of kernels.), and a Deep Learning model. The dataset was split into training and testing sets, with 70% of the data used for training and the remaining 30% for testing. The performance of each model was assessed using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), providing insights into the model's predictive accuracy.

In this study, a GridSearch approach was applied on SVR to find the optimal parameters of the model, including the penalty parameter 'C', the kernel coefficient 'gamma', and the kernel type. Moreover, regarding the deep learning model, several layers, including Dense and Dropout layers was considered. The Dense layers are the core layers in the model where the actual learning happens, while Dropout layers are used to prevent overfitting by randomly dropping some of the neurons during training.

The deep learning model was compiled with the Adam optimizer, which adjusts the learning rate adaptively, and the mean squared error loss function. It was then trained for 100 epochs with a batch size of 32, and a validation split of 0.2, which means 20% of the training data was used as a validation set to monitor the model's performance during training.

4) Evaluation metrics: Post Engagement Rate: The post engagement rate is calculated similarly to the brand engagement rate, but for individual posts. It is a weighted sum of likes, comments, and shares of the post divided by the number of followers at posting. The resulting engagement rates are scaled using MinMaxScaler to ensure comparability. The distribution of Post Engagement Rate can be seen below 10 & 11:

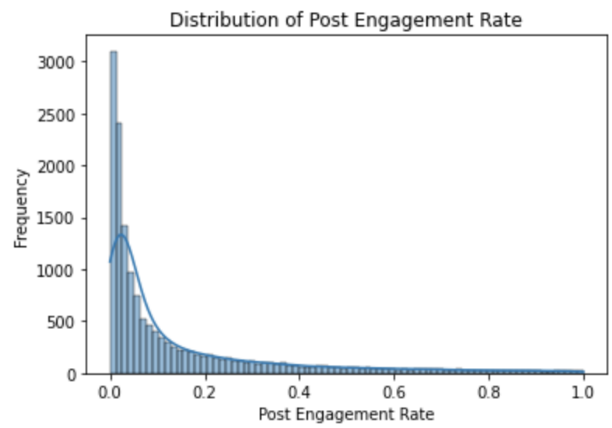

Fig. 10. Posts Engagement Distribution

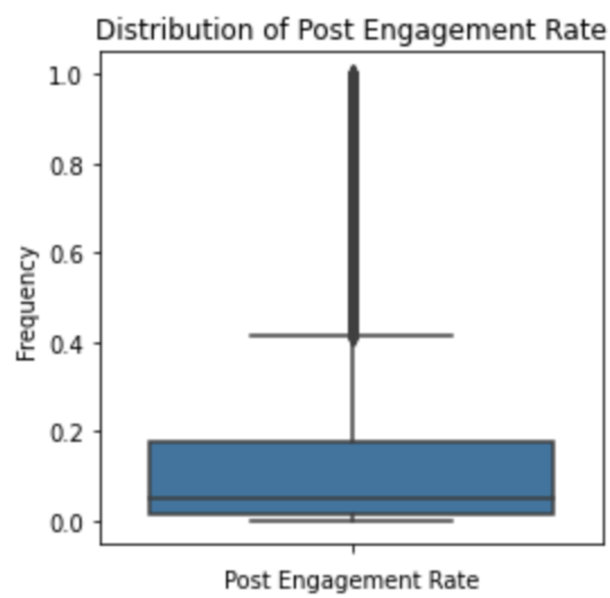

Fig. 11. Posts Engagement Distribution Box_plot

C. Vectorisation Methods: Capturing Semantics

The final section of the methodology details the techniques employed to understand the semantic meaning of post messages and to predict engagement rates based on these semantic features. This approach is aimed at facilitating a comparison with the outcomes from the second research question, which leveraged textual features to predict engagement rates.

Three distinct text vectorization techniques were deployed, each accentuating different aspects of the text:

BERT: The Bidirectional Encoder Representations from Transformers (BERT) model is used for generating sentence embeddings. The specific variant employed here is the pre-trained 'distilbert-base-nli-mean-tokens' model.

FastText: FastText is utilized to form word embeddings. As an extension of the Word2Vec model, FastText treats each word as a composition of character n-grams [19]. This unique aspect of the model equips it with an enhanced ability to generate superior word vectors for rare and misspelled words. The specifications of this model define each vector to have 100 dimensions, considering only words that appear at least three times in the dataset.

Tf-Idf: Term Frequency-Inverse Document Frequency (Tf-Idf) is a numerical statistic that mirrors the significance of a word to a document within a corpus [20]. The technique weighs the frequency of a word in a document (Term Frequency) against its prevalence across all documents (Inverse Document Frequency). The Tf-Idf vectorizer transforms the 'processed_text' into a matrix of Tf-Idf features. It creates a Tf-Idf matrix that includes the top 1000 most frequent words that appear in at least 3 documents but in no more than 90% of the documents. It also accommodates unigrams and bigrams.

Following the text vectorization phase, the resulting vectors from BERT, FastText, and Tf-Idf are leveraged as input in regression models to predict the Post Engagement Rate. For consistency, the same regression models used for feature extraction in the IV-B section are employed in this stage.

V. Experiments & Results

The first experiment of this thesis inquires the influence of airlines' posting frequency strategies on customer engagement rate. As detailed in the methods section, a t-test facilitated the identification of airlines demonstrating significant variations in their posting frequency between the two distinct time periods (normal times and lockdown restriction period) Once these airlines were identified, their engagement and popularity metrics were independently measured for both timelines.

Moreover, the investigation distinguished larger airlines from the smaller ones. This segmentation was promoted by the fact that larger airlines, due to their more extensive brand recognition, might naturally acquire higher engagement and popularity. Segmenting these groups enables an examination into whether it is indeed brand recognition, or the posting frequency that drives engagement rates.

The analysis further includes comparing the brand engagement and popularity scores between the two time periods. This comparison results in finding the percentage change. This percentage change clearly shows if the popularity and engagement have increased or decreased from the first time period to the second. The results of this part of the study are shown below (figures 12 & 13):

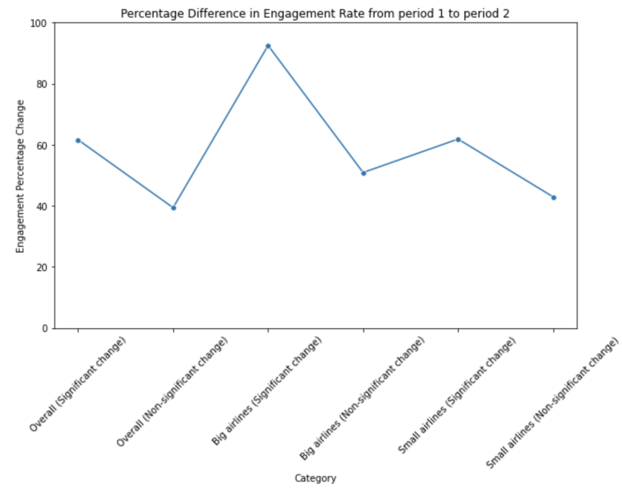

Fig. 12. Percentage change in Brand Engagement period 1_2

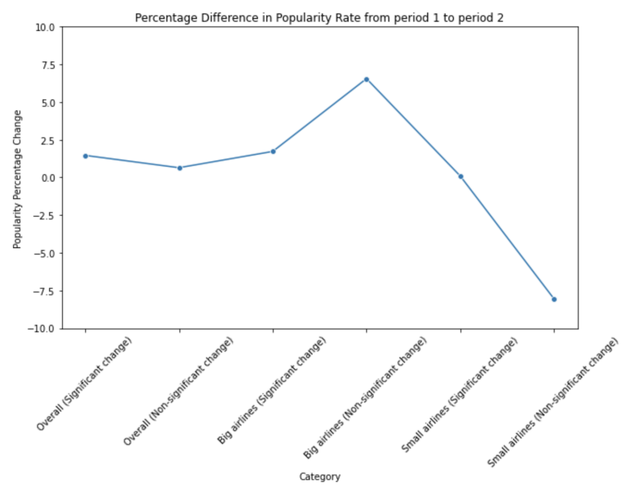

Fig. 13. Percentage change in Brand Popularity period 1_2

It is important to note that the figures12 & 13 are based on the categories that were explained and analysed in the method section IV-A1.

Furthermore, In the methodology explained in Section IV-B, multiple features from the posts were derived. These features were then utilized as inputs to the regression models, to understand their impact on the engagement rate. Initially, all the features were combined and input into the regression models. Then, each feature separately was examined, using each one as an individual input to predict the engagement rate. This process was conducted to determine the most significant feature influencing the engagement rate of posts. The results are displayed in the table below 14 (the lowest values have been marked), also the figures 19, 21 and 20 visualise the errors in the appendices.

The investigation continuous by exploration of semantic features of the posts to predict the Post Engagement Rate. The

	Feature	Model	MSE	MAE	RMSE
0	All Features	Linear Regression	0.0340	0.1256	0.1843
1	All Features	SVR	0.0359	0.1250	0.1895
2	All Features	NN	0.0359	0.1228	0.1895
3	Sentiment	Linear Regression	0.0420	0.1481	0.2050
4	Sentiment	SVR	0.0421	0.1346	0.2052
5	Sentiment	NN	0.0419	0.1506	0.2048
6	Length	Linear Regression	0.0404	0.1466	0.2010
7	Length	SVR	0.0407	0.1333	0.2017
8	Length	NN	0.0402	0.1445	0.2005
9	Time-series	Linear Regression	0.0414	0.1458	0.2034
10	Time-series	SVR	0.0436	0.1379	0.2088
11	Time-series	NN	0.0413	0.1423	0.2031
12	Type	Linear Regression	0.0399	0.1437	0.1999
13	Type	SVR	0.0444	0.1396	0.2108
14	Type	NN	0.0412	0.1458	0.2029
15	Topic	Linear Regression	0.0340	0.1260	0.1845
16	Topic	SVR	0.0363	0.1295	0.1905
17	Topic	NN	0.0360	0.1295	0.1898

Fig. 14. Errors for features

experiments consisted of various vectorization and regression techniques applied on the post messages and extracted features.

The first step of the experiment involves transforming the textual data from the post messages into a numeric form that could be processed by machine learning algorithms. This was achieved by employing three distinct vectorization methods, namely: FastText, TF-IDF and BERT. FastText vectorization was used due to its ability to create vector representations for words while taking into account the internal structure of words. It excels at capturing the semantic and syntactic similarities between words. TF-IDF gives importance to words that are more frequent in a document but less frequent across all documents, thereby giving higher weights to words that are more informative and relevant to the particular context. Finally, BERT vectorization is a more advanced technique that accounts for the context in which a word appears.

Each of these vectorization methods converted the posts into vectors that effectively captured their linguistic properties. Then three distinct regression models were used to predict the Post Engagement Rate based on the vectorized posts, namely: Linear Regression, Support Vector Regression and Neural Networks. Each of these models were trained separately on the vectors produced by each of the vectorization methods.

A comparison between the results of the models with different vectorization can be found in the below table 15 (the lowest values have been marked), and also figures 16, 17 and 18 in the appendices visualise these results:

	Feature	Model	MSE	MAE	RMSE
0	TF-IDF	SVR	0.0304	0.1295	0.1742
1	TF-IDF	NN	0.0317	0.0989	0.1781
2	FastText	Linear Regression	0.0338	0.1274	0.1838
3	FastText	SVR	0.0339	0.1251	0.1840
4	FastText	NN	0.0384	0.1253	0.1958
5	BERT	Linear Regression	0.0318	0.1268	0.1783
6	BERT	SVR	0.0286	0.1195	0.1690
7	BERT	NN	0.0328	0.1095	0.1810

Fig. 15. Errors for vectorisers

VI. Discussion

This section presents the discussion of the conducted experiments and obtained results. In addressing the first research question of this study: "How have the posting frequencies of different airlines changed over time and what impact has this had on their social media engagement and popularity?", several interesting insights were uncovered.

Firstly, it is evident from the results that there is a significant relationship between the posting frequency and brand engagement across the board (figure 12). When considering all airlines, those that changed their posting frequencies significantly saw a considerable increase in brand engagement, from period one to period two, by 60%, while those with non-significant changes observed only a 38% increase. A similar pattern is seen when we separate the airlines into larger and smaller entities. Bigger airlines with significant changes in posting frequency recorded a considerable 95% increase in their engagement rates, as opposed to a lesser 52% rise among their counterparts with non-significant changes. Interestingly, smaller airlines exhibited a similar trend with a 60% increase in engagement for those with significant changes and a 40% increase for those without (figure 12).

When it comes to brand popularity, the results present a slightly different narrative (figure 13). Despite significant changes in posting frequencies, the overall change in brand popularity remained relatively low at 0.7%. The non-significant group saw an even lower increase of 0.1%. A noteworthy point to make here is the impact on the larger airlines. Those with significant changes experienced only a modest increase in popularity (1.7%) while those without, saw a substantial rise of 6.4%. This might suggest that other factors, perhaps the existing brand recognition or content quality, are playing a larger role in brand popularity for these bigger airlines. On the other hand, smaller airlines saw a marginal increase in popularity (0.1%) for those with significant changes, and surprisingly, a decrease of 8.5% for those without significant changes. This suggests that for smaller airlines, maintaining or increasing post frequency could be critical for sustaining brand popularity.

From these results, it is clear that increasing the posting frequency tends to positively influence brand engagement for both big and small airlines. However, the impact of these changes on brand popularity is less straightforward, particularly for smaller airlines where the result is inversely related. Therefore, the airlines that reduced their posting frequency during the COVID lockdown restriction, have lost their customers' engagement in their posts. Moreover, the result suggest that brand popularity might not necessarily be effected by posting frequencies. Further investigation is needed to discover the factors that affect brand popularity.

The second experiment focuses on answering the question: "How can we quantitatively analyze the relationship between specific features of airlines' posts and the level of engagement they receive?". The strategy entailed of extracting features such as sentiment, topics of each sentiment, length, trends and timing and finally the type of each post. Then these features were used as inputs to several machine learning algorithms including Linear Regression, Support Vector Regression (SVR), and Neural Networks (NN) to predict the Post Engagement Rates (in other words finding the most correlated feature to the engagement rate).

Based on the calculated MSE, MAE, and RMSE values, it was found that when all the features were used collectively, a simple model such as Linear Regression gave the lowest error rates among all the three models. However, interesting insights were revealed when these features were used individually. Sentiment, Length, and Time-series performed similarly with Linear Regression and Neural Networks, but Linear Regression yielded slightly higher error rates. Interestingly, when the Topic feature was used individually, both Linear Regression and NN performed remarkably well, suggesting that the Topic of the posts could be a significant factor in driving engagement. Lastly, the Type feature demonstrated similar trends, with slightly higher error rates for NN.

In conclusion, it was observed that a more straightforward model such as Linear Regression outperformed more complex models like Neural Networks (NN) and Support Vector Regression (SVR) in predicting the Post Engagement Rate. Additionally, the Topic of posts yielded the lowest error among all features, with the error rate being comparable to that obtained when using all features collectively. This suggests that when airlines aim to stimulate customer engagement, the posts' topics should be given more consideration. Moreover, the minor difference in error between the topic feature and the composite of all features implies that using just the Topic as a feature could potentially be sufficient in predicting engagement rates.

Lastly, the third experiment investigates the semantic feature of the posts. The question that this experiment tries to answer is "When compared to the features identified in the second part of the study, how effectively can text vectorization techniques, grasp the semantic content and forecast the engagement rate of social media posts?" Therefore, the effectiveness of text vectorization techniques such as BERT, FastText, and TF-IDF were contrasted against the features identified in the second part of our study in understanding the semantic content of airline posts and predicting their engagement rates. It was found that each vectorisation method had unique strengths in capturing various aspects of the text. While TF-IDF effectively emphasized informative words unique to a document, FastText captured semantic and syntactic similarities between words, and BERT accounted for the context of words.

Among the text vectorization techniques, the best performance was achieved by BERT in combination with Support Vector Regression (SVR), yielding an RMSE of 0.169, an MAE of 0.119, and an MSE of 0.0286. The performance of this method is relatively superior compared to the same machine learning model used with other vectorization techniques.

Nevertheless, when we compare these findings to the results of regression models using the extracted features from the second research question, a slightly different picture is portrayed. Here, the combination of the "Topic" feature with Linear Regression proved to be the most effective with an RMSE of 0.184, an MAE of 0.126, and an MSE of 0.034. Interestingly, the performance of the "Topic" feature with Linear Regression was close to that of BERT used with SVR. Moreover, the error rates between these two methods were not significantly different, suggesting that the relatively simple feature extraction method of using the "Topic" feature could be as effective as the advanced BERT vectorization technique in predicting post engagement.

In conclusion, both the features identified in the second research question and the text vectorization techniques demonstrated the ability to predict the post engagement rate effectively. While the "Topic" feature with Linear Regression yielded the lowest error in this case, the BERT vectorization method in combination with SVR also produced comparable results.

VII. Conclusion

This thesis provides a novel perspective on how airlines utilize social media, specifically Facebook, to communicate with their customers. Three key research questions were addressed. Firstly, how the frequency of airlines' posts evolved over time and how these variations influenced their social media engagement and popularity, was examined. Secondly, the correlation between specific features of airlines' Facebook posts and the levels of customer engagement was analysed. Finally, several vectorisation techniques are employed to scrutinize the correlation between the semantics of the posts and the engagement rate, which is then compared to the features extracted from the previous section.

Various findings were concluded of this research. A clear connection was established between an increased frequency of posts and improved customer engagement, particularly during the COVID-19 lockdown period. Nonetheless, the relationship between posting frequency and brand popularity was more complex, indicating the possibility of impact of other aspects such as brand recognition or other factors involving customers' behavior on Social media.

Furthermore, several features of airlines' posts (including sentiment, length, timing, type, content and the topics they discuss) were analysed for their correlation with engagement rates. Linear regression performed better than Support Vector Regression and Neural Networks in predicting Post Engagement Rates. The topic of the posts turned out to be a significant factor in driving engagement, suggesting airlines should consider the topic of their posts carefully to stimulate customer engagement. The findings also suggested that using just the topic as a feature could be sufficient to predict engagement rates.

The last research question investigated the utility of advanced text vectorization methods in understanding the semantic content of airline posts and predicting their engagement rates. While all three vectorization methods (BERT, FastText, and TF-IDF) provided valuable insights, BERT in combination with Support Vector Regression demonstrated superior performance in predicting engagement. Nevertheless, the comparison with the results of regression models using the extracted features showed that the simple feature extraction method using the "Topic" feature could be as effective as the advanced BERT vectorization technique.

Overall, this thesis provides valuable insights into the dynamics of social media engagement in the airline industry. It emphasizes the importance of posting frequency and the topic of posts in driving customer engagement. Furthermore, it demonstrated that simpler models and features can perform comparably to more complex methods in predicting engagement, thus providing practical guidance for airlines in devising their social media strategies.

References

[1] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[2] A. Elemonie, "The role of social media in shaping airline's image," *Jurnal Komunikasi Profesional*, vol. 4, no. 2, 2020.

[3] M. Sigala, "Social media and their transformative role in the airline sector: current issues, trends and challenges," *Journal of Air Transport Studies*, vol. 8, no. 1, pp. 1–22, 2017.

[4] M. L. Moran, G. and D. Johnson, "Message content features and social media engagement: evidence from the media industry," *Journal of Product Brand Management*, vol. 29, pp. 533–545, 2020. [Online]. Available: https://doi.org/10.1108/JPBM-09-2018-2014

[5] M. B. Prados-Peña, E. Crespo-Almendros, and S. Del Barrio-García, "Covid-19 and social media communication strategies: A comparative study of the effectiveness of facebook posts during the lockdown and the "new normal" in the airline industry," *Journal of Air Transport Management*, vol. 103, 2022.

[6] T. Ajiboye, J. T. Harvey, and S. Resnick, "Customer engagement behaviour on social media platforms: A systematic literature review," *Journal of Customer Behaviour*, vol. 18, pp. 239–256, 2019.

[7] Z. Mao, D. Li, Y. Yang, X. Fu, and W. Yang, "Chinese dmos' engagement on global social media: examining post-related factors," *Asia Pacific Journal of Tourism Research*, vol. 25, no. 3, pp. 274–285, 2020.

[8] V. Kulkarni, R. Al-Rfou, B. Perozzi, and S. Skiena, "Statistically significant detection of linguistic change," 2015.

[9] R. Dolan, J. Conduit, J. V. Fahy, and S. Goodman, "Social media engagement behaviour: a uses and gratifications perspective," *Journal of Strategic Marketing*, vol. 24, pp. 261–277, 2015.

[10] P. C. Guerra, W. Meira, and C. Cardie, "Sentiment analysis on evolving social streams: How self-report imbalances can help," in *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 2014, pp. 443–452.

[11] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets," in *Joint Conference on Lexical and Computational Semantics*, 2013, pp. 321–327. [Online]. Available: http://saifmohammad.com/WebDocs/sentimentMKZ.pdf

[12] "Crowdtangle," accessed: 2023-06-23. [Online]. Available: https://www.crowdtangle.com/

[13] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[14] S. learn developers, "Preprocessing data," 2021. [Online]. Available: https://scikit-learn.org/stable/modules/preprocessing.html

[15] M. Honnibal and I. Montani, "spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing," https://spacy.io, 2017.

[16] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[17] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI, 2014.

[18] M. Grootendorst, "Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics," 2020, gitHub repository. [Online]. Available: https://github.com/MaartenGr/BERTopic

[19] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[20] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

Appendices

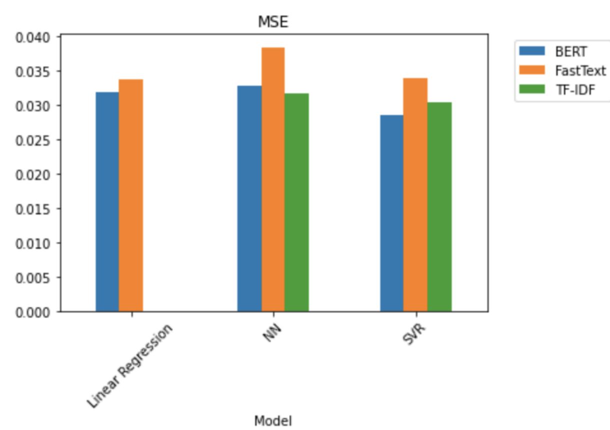

Fig. 16. Mean Square Error

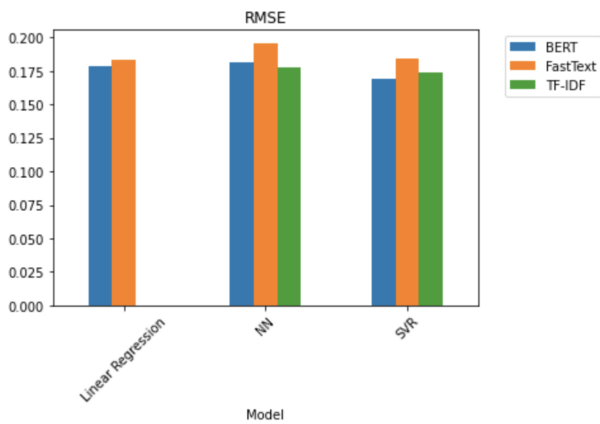

Fig. 17. Root Mean Square Error

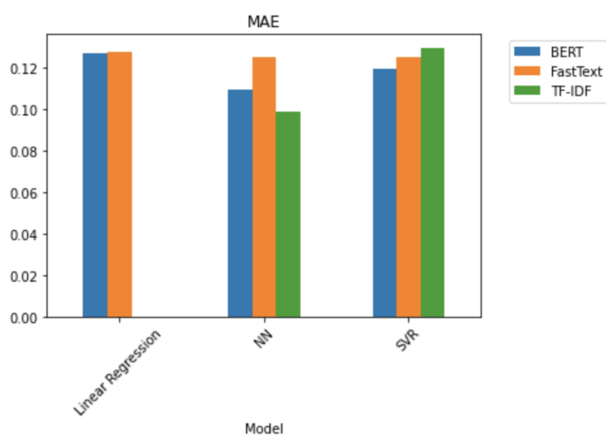

Fig. 18. Mean Absolute Error

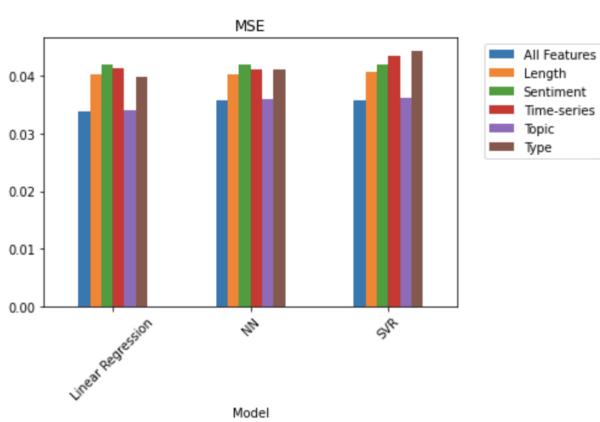

Fig. 19. Mean Square Error for features

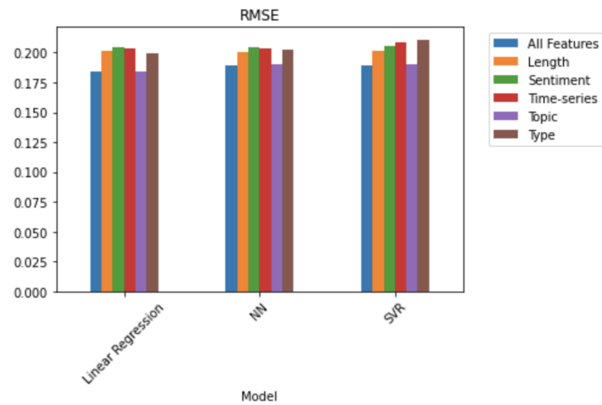

Fig. 20. Root Mean Square Error for features

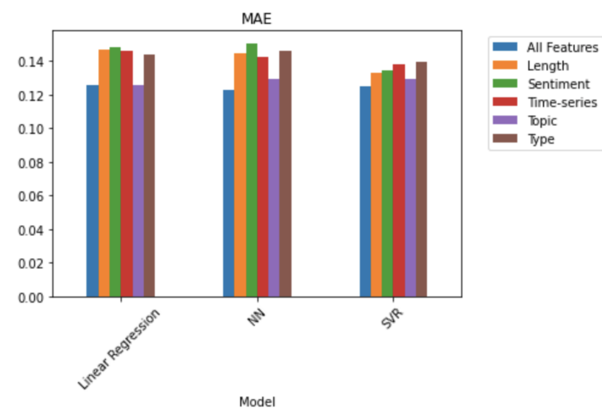

Fig. 21. Mean Absolute Error for features

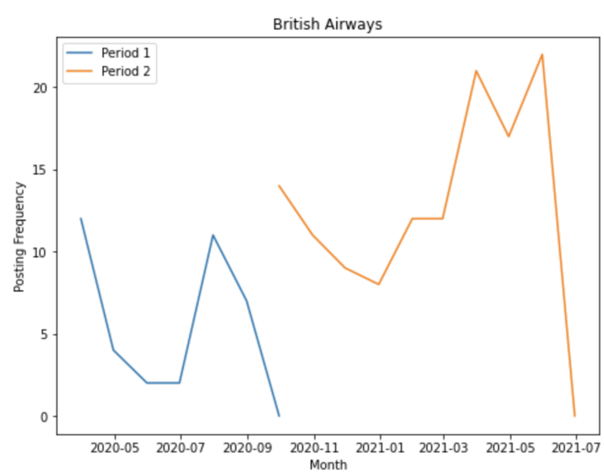

Fig. 22. Different posting frequency airline_1

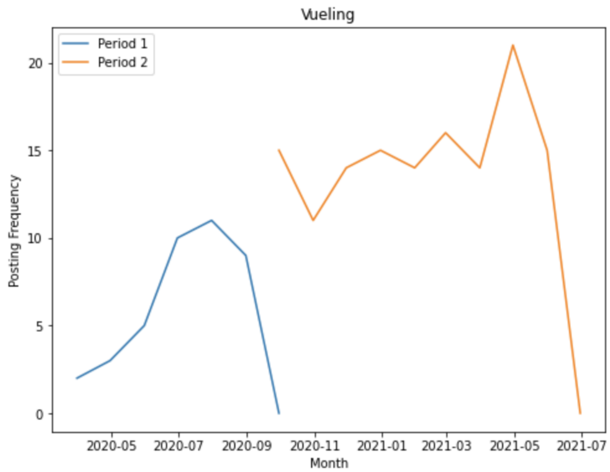

Fig. 23. Different posting frequency airline_2

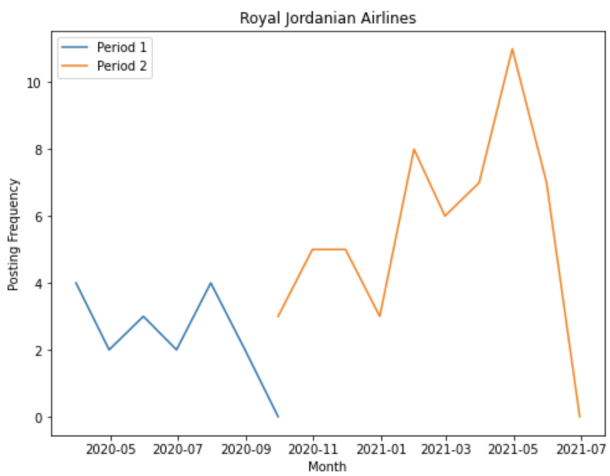

Fig. 24. Different posting frequency airline_3

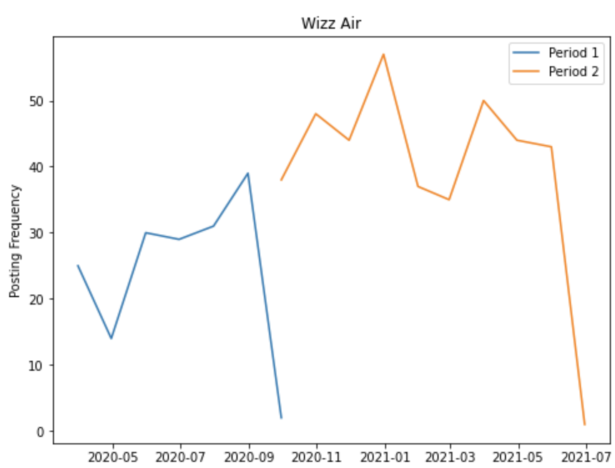

Fig. 25. Different posting frequency airline_4

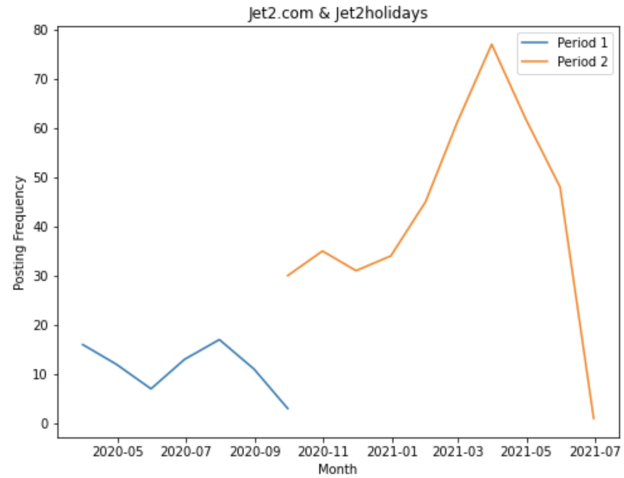

Fig. 26. Different posting frequency airline_5

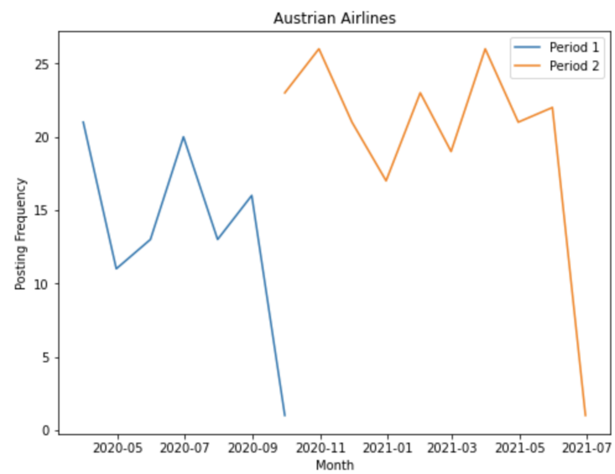

Fig. 27. Different posting frequency airline_6

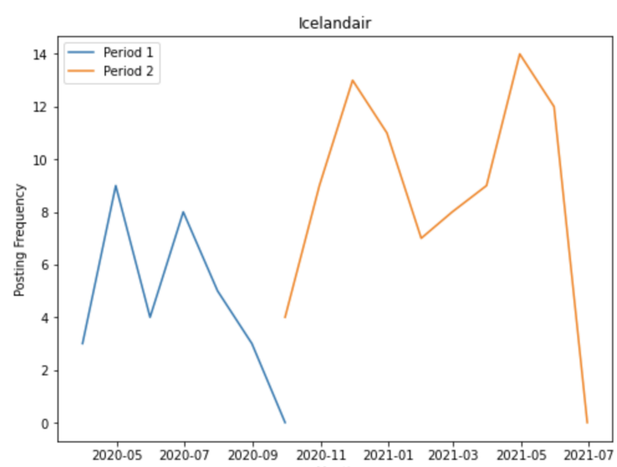

Fig. 28. Different posting frequency airline_7

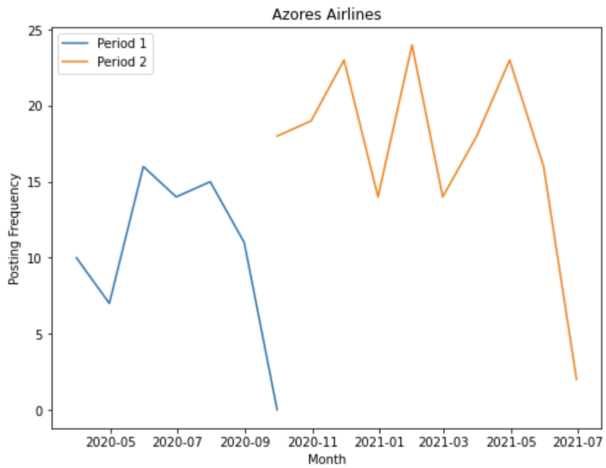

Fig. 29. Different posting frequency airline_8

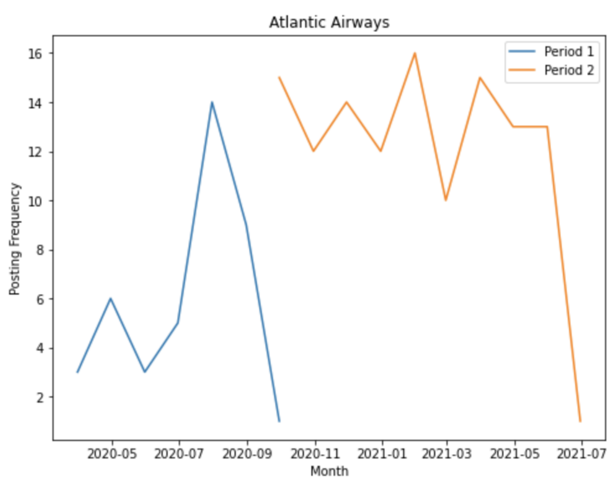

Fig. 30. Different posting frequency airline_9