DAYLIGHT SAVING POLICIES AFFECT ONLINE EMOTIONS

by
MERT ÖZTÜRK

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
July 2022

ABSTRACT

DAYLIGHT SAVING POLICIES AFFECT ONLINE EMOTIONS

MERT ÖZTÜRK

DATA SCIENCE M.S. THESIS, JULY 2022

Thesis Supervisor: Asst. Prof. ONUR VAROL

Keywords: daylight saving time, online emotions, social media, sentiment analysis, time series analysis

The amount of sunlight people are exposed to and the social clock is directly related to Daylight Saving Time (DST) policies. Increasing work efficiency and reducing energy consumption were the original motivations for the policy, but nowadays the effects on human well-being and health are controversial topics. We conducted a social media sentiment analysis to observe the psychological impact of DST on people using a random sample of Turkish tweets between 2011 and 2021. Turkey's exit from the policy in 2016 provides us with an opportunity to compare the periods when DST was applied and when it was not. From this comparison, we find that setting the clocks forward by one hour in the spring (spring-forward) leads to a significant increase in the average sentiment scores of tweets posted over the following 15 evenings. We also found that in DST years, average sentiment scores are significantly higher on spring mornings than on fall mornings. The results of the diurnal and seasonal analyses show that the time shifts due to daylight saving time lead to significant differences in people's moods, especially in the morning and evening hours of the day. Our methodology also helps us point to important societal events in this 10-year observed period. We hope that our findings will lead to the design of better policies for Turkey and improve collective well-being.

ÖZET

YAZ SAATİ UYGULAMASININ DUYGU DURUMUNA ETKİLERİ

MERT ÖZTÜRK

VERİ BİLİMİ YÜKSEK LİSANS TEZİ, TEMMUZ 2022

Tez Danışmanı: Dr. Ogr Uyesi ONUR VAROL

Anahtar Kelimeler: yaz saati uygulaması, çevrimiçi duygudurumu, sosyal medya, duygu durum analizi, zaman serisi analizi

İnsanların maruz kaldığı güneş ışığı miktarı ve sosyal saat, Yaz Saati Uygulaması (DST) politikasıyla doğrudan ilgilidir. Bu politikanın çıkış noktasında İş verimliliğini artırmak ve enerji kullanımını azaltmaktı, ancak günümüzde uygulamanın insan psikolojisi ve sağlığı üzerindeki etkileri odak noktası durumunda. Bu çalışma kapsamında, 2011 ve 2021 yılları arasını kapsayan ve Türkçe Tweetler içeren bir veriseti kullanılarak, DST'nin insanlar üzerindeki psikolojik etkilerini gözlemlemek için bir sosyal medya duygu analizi gerçekleştirildi. Türkiye'nin 2016 yılında uygulamadan ayrılması çıkması, DST'nin uygulandığı ve uygulanmadığı dönemleri karşılaştırabilme avantajı sağlıyor. Bu karşılaştırmaya göre, ilkbaharda saatleri bir saat ileri almak, takip eden 15 günde atılan tweetlerin ortalama duygu puanlarında özellikle akşam saatlerinde önemli bir artışa yol açıyor. Ayrıca, politikanın uygulandığı yıllarda, ilkbahar sabahlarının ortalama duygu-durum skorlarının, sonbahar sabahlarının skorlarına göre anlamlı ölçüde yüksek olduğunu gözlemledik. Gün-içi ve mevsimsel analizlerin sonuçları, DST politikası nedeniyle saatlerdeki kaymaların, özellikle günün sabah ve akşam saatlerinde insanların duygularında anlamlı farklılıklar yarattığını göstermiştir.

ACKNOWLEDGEMENTS

First, I am extremely grateful to my supervisor, Asst. Prof. Onur Varol. This project would not have been possible without his patience, support, and encouragement. His guidance through this process helped me gain a lot of experience for the rest of my life.

Special thanks to all members of VRL Lab for their technical and emotional support during our Master's journey. I sincerely wish the friendships we have gained will be permanent.

I would also like to thank Asst. Prof. Reyyan Yeniterzi for her guidance and valuable comments during Deep Natural Language Processing course.

Many thanks to my family, for their understanding and patience.

Finally, I would like to acknowledge the support of the Scientific and Technological Research Council of Turkey (TÜBİTAK) received under 1002 project (121E013).

To my family...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

In the modern age, the impact of online services on daily life keeps increasing regularly. The quantity of digital information rises rapidly regarding the expansion of the usage of the online tools. This information can be provided by various services such as search engines, e-commerce websites, or social network platforms. Online Social Network services allow individuals to create a public or semi-public profile in a bounded system (boyd & Ellison, 2007). In the 2000s usage of these online platforms was restricted to computers. However, in the last decade, social network services became accessible at any time regarding the spreading of mobile devices.

Today, social network services undeniably influence our feelings, preferences, and social interactions(Ostic, Qalati, Barbosa, Shah, Vela, Herzallah & Liu, 2021). Analysis of the data provided by the social network platforms proposed some innovative approaches for various social and political researchesVarol, Ferrara, Menczer & Flammini (2017). Main advantages of using social media data during these inferences are to increase the number of observation and having an unbiased observation isolated from the observer effect. As a result, making meaningful deductions by using social media data has become one of the top research priorities in Computational Social SciencesTufekci (2014).

The role of social media platforms in sentiment analysis studies has been growing rapidly with the spread of social networks (Kolchyna, Souza, Treleaven & Aste, 2015). The volume and unbiasedness of the social media data make it an effective tool to make inferences about public opinion and sentiment towards different conditions. Likewise, social media data from an extended period carries valuable information about the effects of major public events or policies over timeChoy, Cheong, Laik & Shung (2011). In recent years, studies aiming to detect the effects of significant events on social media users proved that during the periods with higher mood variations there are strong similarities between mood surveys and social media sentiments (Pellert, Metzler, Matzenberger & García, 2021).

1.1 Daylight Saving Time Policy

Daylight Saving Time (DST) is a policy that advances clocks (generally by 1 hour) between March and October. Setting clocks forward in the springs is called spring-forward while the reverse of this movement in the falls is called fall-backward. The idea of the procedure arose in the late 19th century and became widespread in less than a half-decade. The first nationwide applications of DST have been seen in Europe due to energy-saving concerns.

In the early 1900s, more daylight meant less usage of artificial lights. However, after the worldwide spread of electronic devices other than electric bulbs, energy consumption became less related to daylight. Due to studies in the last decades, effects of DST on energy consumption have been becoming less significant (Havranek, Herman & Irsova, 2018).

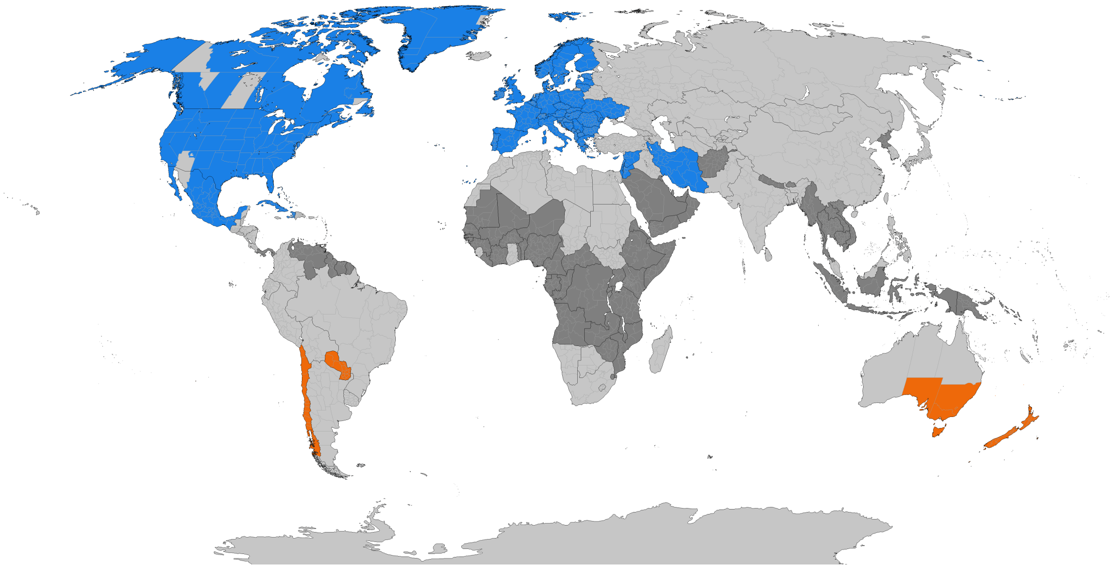

Figure 1.1 **Countries applying DST.** Countries represented by different colors according to policy implementation. Blue :Observes DST around the Northern Hemisphere summer, Orange:Observes DST around the Southern Hemisphere summer, Light Gray:Formerly observed DST, Dark Gray: Never observed DST **source**:https://en.wikipedia.org/wiki/Daylight$_{s}aving_{t}ime_{b}y_{c}ountry$

Today, 73 countries apply Daylight Saving Time (DST) at least one city or state of the country. By this expansion, DST became one of the most widespread and controversial policies in the World. Fig 1.1 displays these countries with blue colors. Turkey applied the policy at various intervals between 1940 and 2016. In 2016, Turkey abolished the policy and began to use UTC+3 timezone permanently.

1.2 Motivation and Resarch Questions

This research aims to observe the within-day and seasonal effects of DST, on the sentiments of the social media users in Turkey. Ease of accessibility and worldwide usage make Twitter a plausible social media platform for this study. Twitter is one of the leading social media platforms with over 320 million users around the World, and Turkey is the 2nd country in Europe and 7th country in the World with more than 16 million accounts (Statista, 2021). Each day, there are half-billion tweets sent on average (Sayce, 2020), and these tweets provide valuable traces about the psychical and mental conditions of the users (Tong, Zhang, Zhang, Sadka, Li & Zhou, 2019).

Previous studies about the effects of DST on psychology are generally based on various resources such as medical records or surveys(Hansen BT, 2017). The main motivation for using online data during this study is to collect more data with less human effort and to obtain a more unbiased sentiment estimate without an observer effect. The dataset used during the study includes about 1.5 billion Turkish tweets between 2011 and 2021. One way to detect to sentiment classes of these tweets to train a Natural Language Processing (NLP) model for the sentiment classification task. A sub-sample of our dataset was annotated in 3 different sentiment classes (negative, neutral, and positive) and used in the task-specific fine-tuning process of Bidirectional Encoder Representations from Transformers (BERT)Devlin (2018).

Sentiment scores obtained from this model were used to calculate a mean estimate for sentiments of one-hour periods. These mean estimates of hourly sentiments provide valuable information on time-dependent changes in Twitter sentiment between 2011 and 2021. During this period, Turkey applied DST policy until 2016 and started to utilize the permanent UTC+3 timezone in 2017. This situation gives us a chance to compare the sentiment changes during the years with DST and without DST. This study proposes 4 main hypothesis about effects of DST on Twitter sentiment.

- How does DST affect within-day sentiments?
- How does DST affect seasonal sentiments?
- How does social clock affect within-day sentiments?
- Does societal sentiment reflect important societal events?

The research question about the within-day impacts examines the distribution of the sentiments by hours of the day for two different groups. These groups are periods before and after DST. Moreover, within the scope of this research question years with and without DST and spring-forward and fall-backward periods were also compared. The second research question focuses on the hourly distribution of the sentiments for different seasons of the year for the year with DST and without DST. The question about the social clock compares the effects of the amount of daylight and the daily social clock on social media sentiment. The social clock is the term that defines the time of daily habits of the people such as, work shifts, school start times. The last question examines the effects of major historic events such as terrorist attacks, elections, national holidays, or religious festivals on Twitter sentiment. This question was supported by examining the change in the sentiment scores, as well as detecting the top 10,000 daily popular words and the popularity change of these events over time.

1.3 General Flow of the Thesis.

A review of the previous studies that focus on the effects of DST will be given in Chapter 2. Chapter 3 describes the dataset and clarifies the methods applied to obtain time-dependent sentiment estimates. Chapter 4 proposes the research questions and findings of this study. In chapter 5, conclusions and implications of DST policy.

2. Literature Review

Today, 73 countries apply Daylight Saving Time (DST) at least one city or state of the country 2.1. This number drives DST, one of the most widespread and controversial interventions in the World. While discussing the effects of the DST policy, we must recognize the critical role of daylight on people's circadian rhythms (Roenneberg, Wirz-Justice, Skene, Ancoli-Israel, Wright, Dijk, Zee, Gorman, Winnebeck & Klerman, 2019). Studies based on Twitter activities have shown that diurnal sentiment and seasonal sentiments got affected by several factors such as weekend effects or seasonal effects like daylight length and people are happier on weekends but the morning peak in the positive effects is delayed by 2 hours on those days (Golder & Macy, 2011). Human circadian rhythm also responds to time shifts caused by DST both in the mornings and evenings (Rishi, Ahmed, Perez, Berneking, Dombrowsky, Flynn-Evans, Santiago, Sullivan, Upender, Yuen, Abbasi-Feinberg, Aurora, Carden, Kirsch, Kristo, Malhotra, Martin, Olson, Ramar, Rosen, Rowley, Shelgikar & Gurubhagavatula, 2020). For this reason, in addition to its effects on energy consumption, studies on Daylight Saving Time (DST), also focused on various impacts of the policy on health, and psychology.

2.1 Effects of DST on Studies on Energy Consumption

Research in 2018 focused on the impact of DST on energy consumption (Havranek et al., 2018). In this research, 162 estimates from 44 studies including government papers, energy company reports, and articles were assembled. This literature implies when DST is applied, the estimated ratio of electricity saving equals 0.34%. However, this estimated average impact is a biased parameter of the actual impact of DST on energy consumption. The distribution of the estimates in energy economics is truncated due to publication bias and study design. When the research only focused on the more prestigious estimates with better data and methodology, the results did not show any evidence of electricity savings due to DST. On the other hand, they have found that estimated electricity-saving increases in the countries with higher latitudes.

According to a study conducted in Turkey in 2010, it is estimated that 10% savings will be obtained from the electricity used for lighting in the residences, in a situation where the clocks are advanced by 30 minutes (Karasu, 2010). This means 0.7% savings in electricity used throughout the country.

2.2 Effects of DST on Health and Psychology

DST causes seasonal variations in the sleep cycle and, labor cycles in industrialized countries (Martín-Olalla, 2019). This research also focused on how the effects of DST vary over the countries with different latitudes. To measure how the activities of the people changes over time, Time Use Surveys (TUS) are used as a tool. The results from these large-scale studies (N $\sim$ 10,000) showed that the labor cycle is equally distributed through seasons.

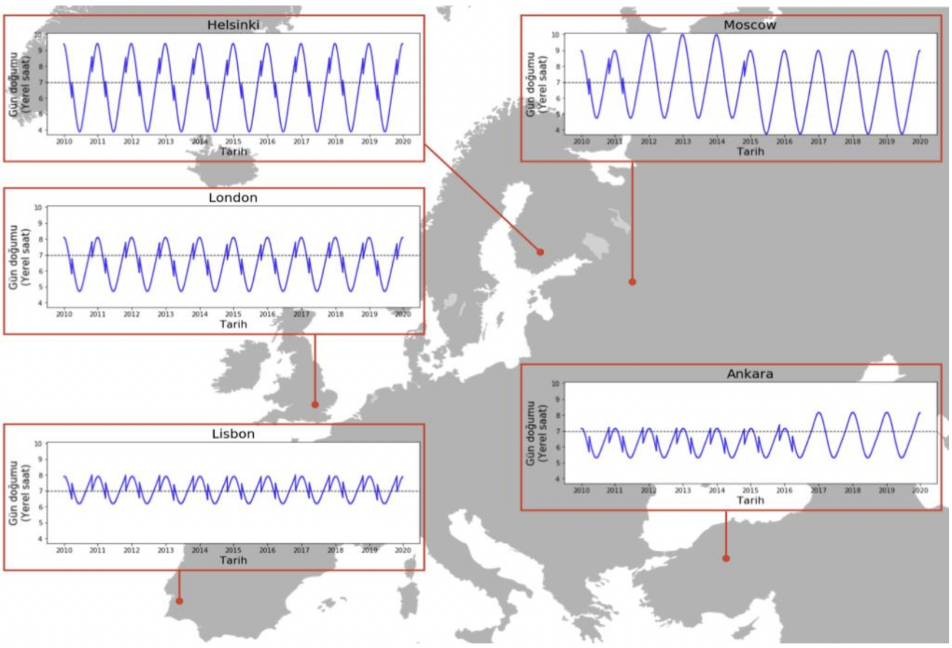

Figure 2.1 **Sunrise time in Europe.** Sunrise times represented with darkblue lines for 5 different countries from different latitudes. Shifts in the graphs represent the DST days.

Fig 2.1 visualizes the the sunrise times of countries at different latitudes. Under free preferences on the weekends, it is observed that the sleep cycle is more prone to seasonal deviations, especially in lower latitudes. The sleep/wake cycles in Great Britain and France show fewer excursions than the sleep/wake cycle in Spain and Italy. Thus, the effects of the DST on daily cycles vary over countries located at different latitudes. However, these deviations are still linked to solar activity and the activities at noon are not misaligned by the DST intervention.

In (Hansen BT, 2017) to observe the psychological effects of DST, the researchers focused on the number of unipolar depressive episodes. The nationwide dataset was collected from Danish Psychiatric Central Research Register from 1995 to 2012. Time series intervention analyses were based on 185,419 hospital contacts for unipolar depression. Due to results, during the fall-backward period, the incidence rate of unipolar depressive episodes diagnoses increased 11 percent (95% CI = 7%, 15%). This increase in the incidence rate fadeout over approximately ten weeks.

2.3 Social Media Sentiment Analysis

Social media data carries huge information about public opinion and people can use social media as a tool to both express their positive and negative emotionsAgarwal, Xie, Vovsha, Rambow & Passonneau (2011). When the time-dependent changes of these sentiments are examined, duration of the positive and negative online emotions varies from each other (Fan, Varol, Varamesh, Barron, van de Leemput, Scheffer & Bollen, 2018). Along with sentiment analysis, there are various studies and products that focused on social-media specific NLP tasks such as emoji based sentiment prediction (Choudhary, Singh, Bindlish & Shrivastava, 2018) and offensive language detection(Camacho-Collados, Rezaee, Riahi, Ushio, Loureiro, Antypas, Boisson, Espinosa-Anke, Liu, Martínez-Cámara, Medina, Buhrmann, Neves & Barbieri, 2022) (Jahan & Oussalah, 2021). The studies in the field of social media sentiment analysis aim to observe sentiments and feelings of the people on different domains.

Political tweets can form one of these domains. There are studies that examine the sub-branches of tweets about political content such as Black Lives Matter and Stop Asian Hate movements and the relationships between these contents (Xin Tong, 2022). Some of the studies on political tweets have proposed their own hybrid models, due to the challenging nature of social media NLP tasks (Ozdemir & Yeniterzi,

2020) and because of the complexity of the language they focused on (Safaya, Abdullatif & Yuret, 2020). Politicians also constitute the area of interest in studies in this field. A study in 2016 focused on the main actors of the political discussions and tweets of former U.S President Barack Obama and U.S 50 Governors (Yang, Chen, Maity & Ferrara, 2016). They discover Republican and Democrat Governors are similarly active on Twitter but use different styles of communication.

One of the most popular domains for sentiment analysis is financial market prediction. In a study on that domain, researchers aimed to explore sentiment metrics that are correlated with the price movements of Bitcoin (Raheman, Kolonin, Fridkins, Ansari & Vishwas, 2022). They tried 15 BERT-based approaches for the sentiment analysis of the Tweets and Reddit posts and fine-tuned those models with texts including specific words from cryptocurrency terminology and jargon. According to the results, an interpretable NLP model named 'Aigents' has a 57% correlation with the price movements of Bitcoin.

2.4 Relationship Between Sleep Time and Social Media

An intervention can affect society in many different ways. It is possible to measure those effects by analyzing the social media activities of the users (Linnell, Alshaabi, McAndrew, Lim, Dodds & Danforth, 2020). This method both saves the human effort for data collection and provides a larger amount of observation compared to surveys or medical records. Past works in this area showed there is a correlation between low activity on Twitter and sleep time (Leypunskiy, Kıcıman, Shah, Walch, Rzhetsky, Dinner & Rust, 2018).

By leveraging the correlation between sleep time and Twitter activity, it can be also possible to detect the sleep time changes in the Daylight Saving Time (DST) periods (Linnell et al., 2020). In this study, researchers proposed estimated sleep loss using the Twitter activity of two different groups. Those groups include the period before spring forward (BSF) and the period after spring forward (SF) between the years 2011 and 2014. For these two different groups compared with both the tweet includes some keywords (breakfast, lunch, dinner) and the normalized count of the random tweets posted on Sunday and Monday. The results showed that peak twitter activity occurs approximately 45 minutes later on Sunday evening following spring forward. The meal-related tweets also show a small forward shift after the spring forward. While the correlation between screen time and lack of sleep is considered, it is possible to say that the time shift in Twitter activity can be the indicator of the effect of DST on the sleep cycle.

3. Methods

3.1 Dataset details and statistics

The dataset was collected by Twitter's streaming API which provides randomly selected 10 percent of the public data stream. During the streaming, tweets have been collected based on the "lang" parameter that indicates the user and tweet language. Since the collection criteria of the tweets are the language parameter, the content of these tweets is not limited to specific queries or areas.

Turkish Twitter Dataset includes over 1.5 Billion Turkish tweets from various fields such as politics, economics, sports, technology, and daily life. This diversity provides more unbiased information to make inferences about the hourly average sentiments of society. Each observation in the data consists of tweet metadata that has dozens of features. During this study, only text and creation time features were processed. The final version of the dataset includes text variables of the tweets from 2011 to 2021.

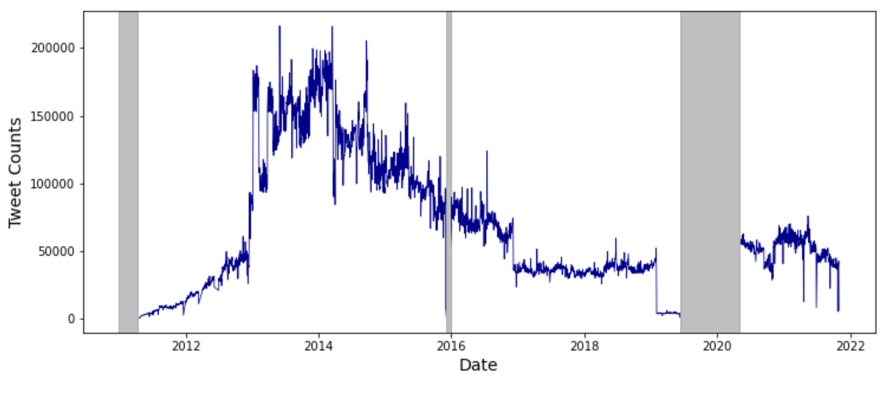

Figure 3.1 **Daily tweet counts.** Tweets counts are presented for each day and the missing data are highlighted in a gray region.

Fig 3.1 displays the distribution of tweet count per day. The tweets created between 2013 and 2015, constitute the majority of the dataset. There are some gap periods in our timeline, due to unintentional interruptions in the scrapping system. In the periods of April 2019-October 2011, December 2015, and June 2019-March 2020 there is no observation in the dataset.

3.2 Sentiment Analysis and Modeling

3.2.1 Sentiment Analysis

Sentiment analysis is one of the major tasks of Natural Language Processing. Sentiment analysis aims to classify texts according to their polarities or opinions towards specific or generic topics (Song, Wang, Jiang, Liu & Rao, 2019). In this study, a Turkish sentiment analysis tool had to be used for tweet classification. The most challenging factor of the sentiment classification with social media data is the informal and daily languages of the social media users.

The majority of the tweets include noisy texts with spelling errors, repetition of the characters, emoticons, wrong usage of suffixes, hashtags, weblinks, etc. Besides, being a morphologically complex language (Yeniterzi & Oflazer, 2010) makes Turkish a challenging case for text classification. As a result of all these, using a powerful language model becomes a necessity for Turkish text classification. In recent years, the models based on Transformers architecture have played crucial roles in NLP and given more accurate results on text classification tasks than its processors.

3.2.2 BERT

In recent years the models based on Transformers architecture play crucial roles in NLP and gives more accurate results on text classification tasks than its processors. In 2019, a new language representation model Bidirectional Encoder Representations from Transformers (BERT), was published. BERT pre-train with unlabeled text by conditioning both left and right context in both layers. (Devlin, 2018). Along the pre-training process, BERT uses a Masked Language Model (MLM). MLM randomly masks some of the tokens and aims to predict the actual word with cross-entropy loss. In addition to the MLM process, a next sentence prediction task is also used to pre-train text-pair representations.

BERT model includes 12 Transformer blocks, 12 self-attention heads, and a hidden size of 768. The embeddings obtained from different layers of the pre-training process include valuable semantic and syntactic information for different NLP tasks such as text classification, question answering, etc. In this study, an open-source pre-trained BERT model was used. This model (Schweter, 2020) is pre-trained with Turkish OSCAR Corpus. The final training corpus includes more than 40 billion tokens with 128 thousand vocabulary sizes.

3.2.3 Fine-tuned BERT for Sentiment Classification

This unsupervised pre-training is followed by a task-specific fine-tuning process (Chi, Qiu, Xu & Huang, 2019). In this study, the pre-training model requires data labeled for sentiment classification. The Turkish tweet dataset includes text variables of the tweets from 2011 to 2021. From each month of the dataset 100 tweets were obtained by random sampling tools prepared for the annotation stage. After the invalid tweets, most of which were posted by bot accounts were eliminated, 2612 were transferred to the fine-tuning process.

Each tweet was annotated according to the sentiment of the user language. For a tweet, there are 3 possible classes: negative, positive, and neutral. During the annotation process, a specific guideline was followed to minimize the differences caused by the semantic differences.

Negative Sentiment : The negative sentiment category includes strong negative cases such as insults, swearwords, and harsh criticisms. Moreover, the tweets with negative opinions and pessimistic or depressive feelings belong to that group.

Positive Sentiment : The positive sentiment category includes tweets that include strong positive opinions about a situation or an experience, greetings, blessings, or praise belonging to the positive group. Apart from that, the tweets about love and emotional relations are classified under this group.

Neutral : The neutral group is the majority group in the annotated data. This group mainly includes the tweets that do not show a strong personal opinion in other words objective tweets. Informative but not evaluative tweets, the news or non-sarcastic question sentences belong to that group. Also, the tweets which include multiple emotions at different parts of the sentence and do not show any clear average sentiment are classified under the neutral group.

Tweets and Their Classes	
Tweet	Class
`Oyle güzel, ¨oyle tatlı anlatamam.. :))`	Positive
`SIZE DE OLUYOR MU BU??"`	Neutral
`bosuna yorulmussun bisiye de benzememis zaten`	Negative

Table 3.1 **Tweet annotation** Tweets were annotated under 3 different classes: positive, neutral, negative

The most indecisive part of the annotation process was discriminating between the neutral-positive or neutral-negative groups. Even for a human to classify a sarcastic question or a sentence that includes multiple emotions can be ambiguous. Also, the news that includes some negative context but does not indicate any personal opinion is another problematic sentiment classification task for a language model. The final dataset contains 2612 manually annotated Turkish tweets in 3 different categories: 757 negative, 569 positive, and 1286 neutral.

During the fine-tuning process, we utilized this fine-grained Turkish sentiment dataset. This model was trained with a learning rate of 5e-5, batch size of 32, for 4 epochs. F1, Recall, and precision scores were saved to evaluate the class-specific performance metrics of the model.

Performance Metrics of Fine-tuned Turkish BERT			
Class	F1	Recall	Precision
`Negative`	0.76	0.74	0.79
`Neutral`	0.76	0.79	0.77
`Positive`	0.80	0.81	0.80

Table 3.2 **Comparison of the Macro Avg. Metrics** Class based performance metrics are balanced

Table 3.2 shows that the performance metrics of the model are balancedly distributed over different classes. Although the data used during the fine-tuning process is imbalanced, our model does not tend to classify an observation heavily on the majority neutral class.

3.3 Hourly Grouping and Bootstrapping

The main goal of this study is to detect the time-dependent changes in the sentiment score. For each of the tweets in the dataset, there are 3 possible classes but these sentiment classes are not individually operable for this analysis. The sentiment classes obtained from the BERT model were divided into intervals according to the creation time of the tweet they belonged to. These periods start at the top of the hours and last until the next hour. Every one-hour interval includes three different values 0, 1, and 2 and these values correspond to negative, neutral, and positive tweets respectively.

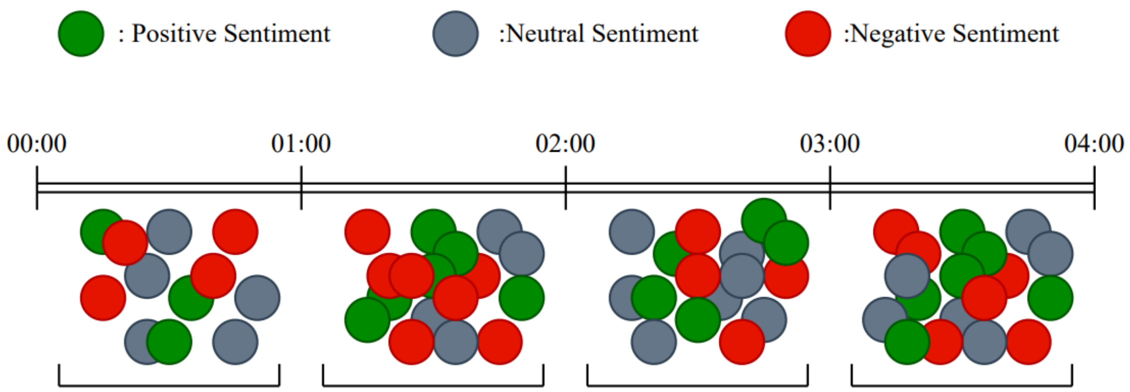

Figure 3.2 **Hourly grouping process** Sentiment groups were represented in 3 different colors and grouped by hours

Fig 3.2 displays the hourly grouping process of the sentiment scores of the tweets. For each of these one-hour intervals, an estimated hourly sentiment score is needed to be calculated. To calculate this estimated sentiment for each of these groups a random sampling technique named bootstrapping was applied. Under the bootstrap sampling technique, 1000 samples were created for each interval. Each of these samples has N (number of tweets posted on that interval) randomly picked sentiment scores, and in every step, each score has an equal chance of being selected.

Fig 3.3 demonstrates the bootstrapping and the new samples created for a group by this method. The grand mean of the averages of these 1000 samples gave an estimated sentiment score of an hour. This process was repeated for a total of 81,467 one-hour intervals between 2011, and 2021. This method provides a credible metric to examine the time-dependent behaviors of the Turkish tweet sentiments.

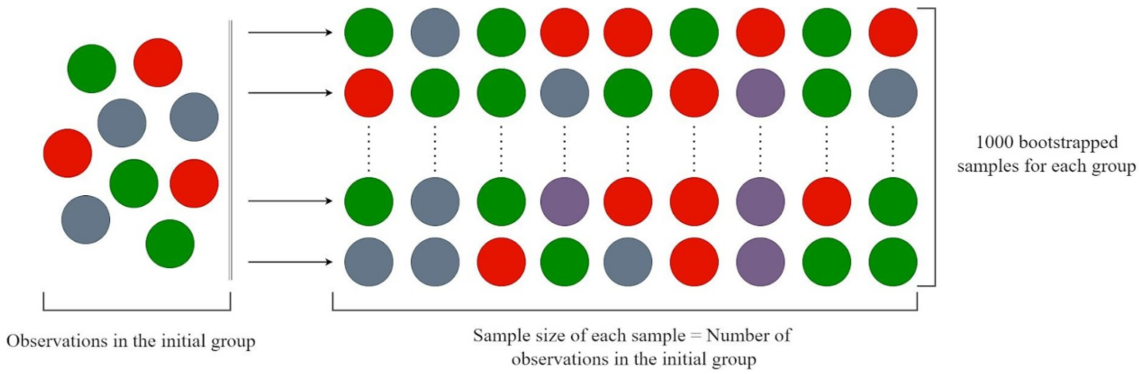

Figure 3.3 **Procedure of Bootstrap Sampling** Bootstrap sampling were applied for each hourly group between 2011 and 2021

3.4 Time-series Extraction and Detection of External Events

By merging consecutive estimates of hourly sentiment, a time-series graph that displays the change in average sentiment from 2011 to 2021 was created. Instead of evaluating the whole ten years period at once, yearly, monthly, and daily time-series graphs will provide valuable information about particular time periods.

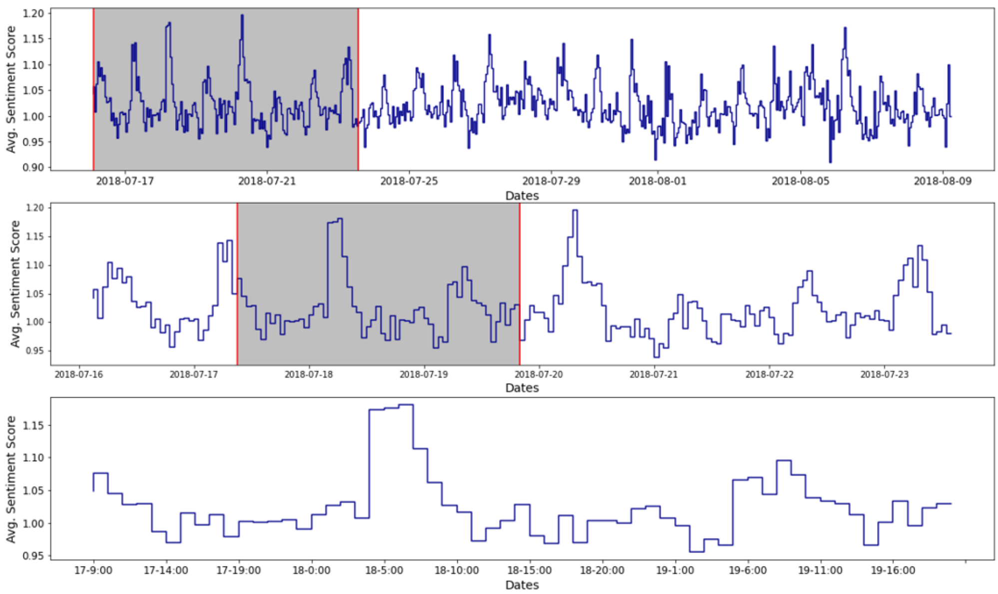

Figure 3.4 **Change of Average Sentiment Over Time** Time dependent changes of the estimated sentiment for 1 month, 1 week and 2 days

Fig 3.4 shows the change in average sentiment over different periods of July 2018. In the 3^{rd} figure, that focus on 48 hours, it is easier to observe the horizontal lines representing the estimated hourly sentiments. These time-series graphs also display the daily and weekly patterns of the sentiment scores. Visualizing within-

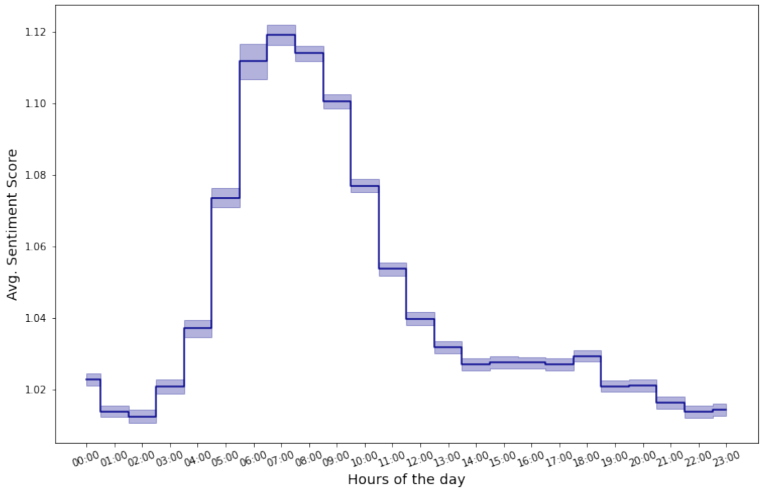

Figure 3.5 **Distribution of Avg. Sentiment Scores by hours of the day**
Distribution of the within-day sentiments

day hourly sentiment averages and within-week daily sentiment averages provides the characteristics of these patterns.

Fig 3.5 displays the within-day patterns that average sentiment scores follow by grouping the hourly estimated scores and calculating the average of them. This graph shows the hourly sentiment scores peak early in the morning between 06:00 and 09:00 A.M. However, the late at night between 01:00 and 03:00 A.M average of the hourly sentiments dropped.

Fig 3.6 shows the averages of the hourly sentiment scores according to days of the week. Considering the within-week daily averages, the sentiment scores reflect the effect of the pre and post-weekend periods and peak on Fridays and drop on Mondays and Tuesdays.

3.4.1 Detecting Major External Events

Although the time-series graphs of the sentiment scores follow daily and weekly patterns, there are some points that lead to anomalies in this order. The sharp increase or decrease in time-series graphs that display the change in hourly estimated sentiment can be caused by various reasons. One of the main motivations behind detecting the external effects that cause fluctuations in the time-series sentiment graph was trying to isolate the effects of DST from various public events. While

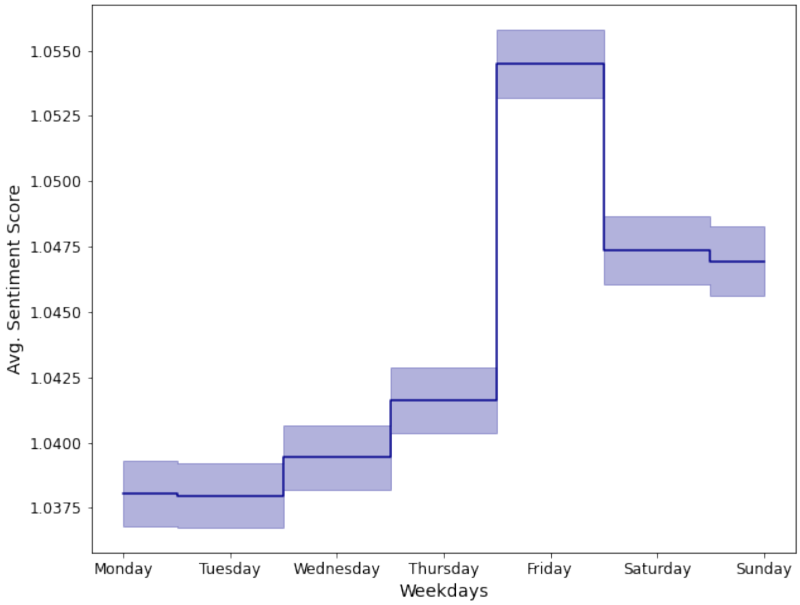

Figure 3.6 **Distribution of Avg. Sentiment Scores by hours of the day**
Distribution of the within-week sentiments

determining the causes of these fluctuations, the most frequently used 10,000 words of each day were used.

During this process, it was aimed to label the words with higher ranking variations as external event words and new hourly sentiment estimates calculated using tweets that do not contain these external event words. However, on the days when external events occurred, no significant difference could be obtained between these new sentiment estimates and the initial ones.

The explanation behind this is that when there is a major public event, not only the tweets about that event but also other tweets can be positively or negatively influenced by that event. After reaching this conclusion, instead of removing these public events from our dataset, we focused on a new research question that examines the effects of some important major historic events that have taken place in Turkey in the last 10 years.

4. Results

4.1 Specifying DST Periods

4.1.1 DST in Turkey

Although the Turkish Twitter Dataset includes tweets between 2011 and 2021, there are some gap periods as Fig 3.1 displays. Moreover, due to scrapping errors caused by rate limits of Twitter API, for the years after 2020, only a few tweets were scraped between 05:00 and 06:00. As a result of all these, during analysis regarding research questions, the period between 2012 and 2020 was used.

Spring-Forward		Fall-Backward	
Date	Rank	Date	Rank
2011/03/28	87	2011/10/30	303
2012/03/25	85	2012/10/28	302
2013/03/31	90	2013/10/27	300
2014/03/31	90	2014/10/26	299
2015/03/29	88	2015/11/08	312
2016/03/27	87	2016/09/08	252
UTC +3 after 2016			

Table 4.1 **DST Periods in Turkey.** The rank of the DST date in the year between the years 2011, and 2016

Table 4.1 shows the days of DST application in Turkey between 2011 and 2016. As mentioned in the previous parts of this study, Turkey applied the DST policy until 2016. For the years between 2012 and 2016 Turkey made spring-forward and fall-backward shifts generally on the night connecting Saturday to Sunday. In 2011 due to the Transition to Higher Education Examination (YGS) and 2014, due to local elections in Turkey spring-forward shifted from Sunday to Monday.

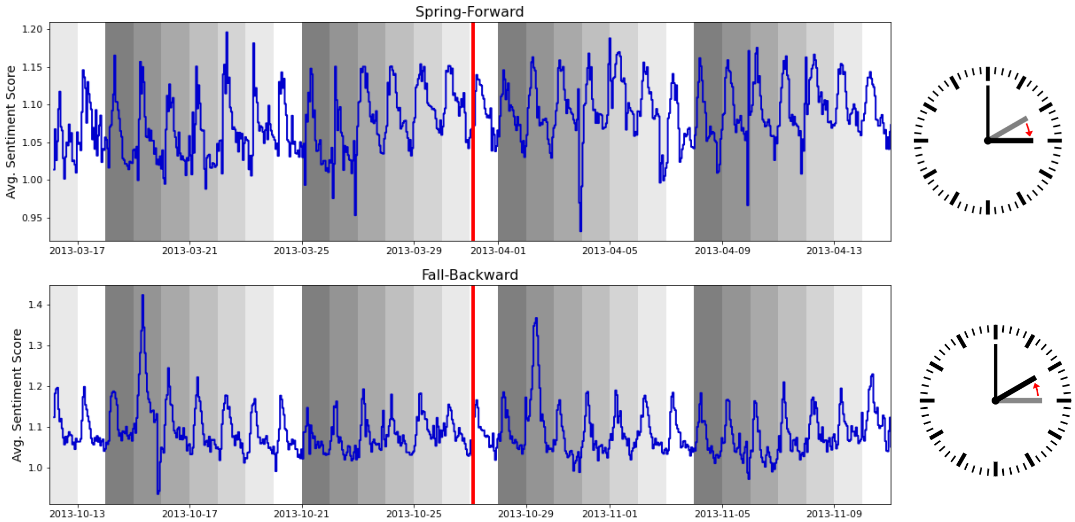

Figure 4.1 **Spring-forward and fall-backward periods of 2013.** Estimated sentiment in 1 month spring-forward and fall-backward periods in 2013 displayed with dark blue line. Background colors represents weekdays.

Figure 4.1 displays the change in average sentiment score during 15 days periods before and after the spring-forward and fall-backward periods of 2013. The red vertical line represents the moment that 1-hour time shifts have occurred. Apart from pre-post DST periods the graph also shows the daily seasonality trend that peaks early in the morning and bottoms after midnight. This recurrent circadian rhythm shows that on average people tend to share more positive tweets in the mornings while negative tweets are late at night. To observe the weekly trends of the sentiment scores, the background of the figure symbolizes the days of the week. There 7 different colors on the background and the days from Monday to Sunday are represented by colors going from white to dark gray.

As mentioned in the Section 3.4.1 these monthly average sentiment graphs tend to get influenced by the external events in Turkey. One of the main challenging factors of this study is to isolate the hourly, daily, and monthly DST period analyses from the impact of these external events. To minimize the effect of the external events rather than directly focusing on the monthly individual DST periods estimates from all spring-forward periods between 2012 and 2016 are included in the analysis. By taking into consideration all estimates from the spring-forward and fall-backward periods of these years new graphs that display the hourly distribution of the estimated sentiments were created.

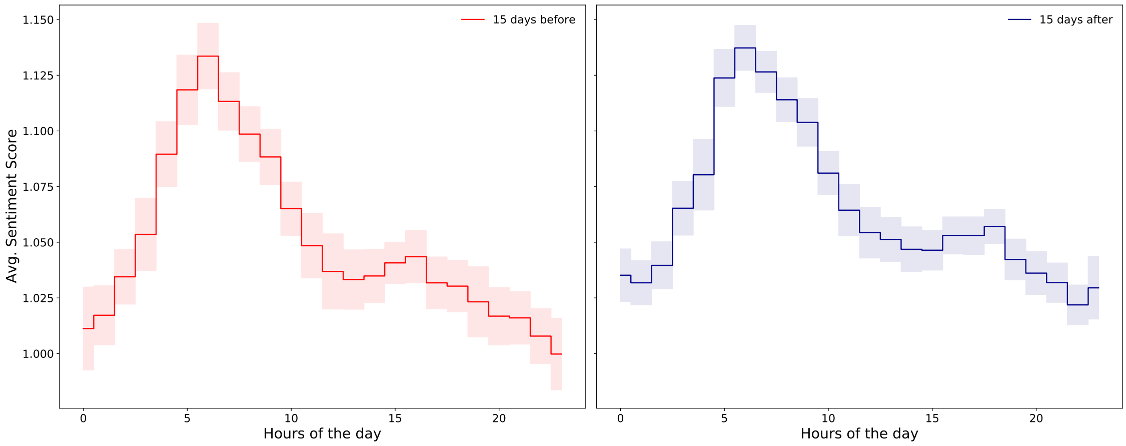

Figure 4.2 **Within-day comparison of pre and post DST periods.** Within-day sentiment distribution of the 15 days periods before and spring-forward

Fig 4.2 shows the hourly distribution of the sentiment scores for the periods 15 days before and after DST. The red horizontal line on the graph on the left-hand side shows the averages of hourly estimated sentiment scores from the pre-DST periods the years between 2012, and 2016. The blue horizontal line on the graph on the right-hand side represents the averages of hourly estimated sentiment scores from the post-DST periods. Similar to Fig 3.5 in both periods (15 days before and after DST) the within-day sentiments peaked early in the morning and dropped late at night.

4.2 Research Questions

4.2.1 How does DST affect within-day sentiments?

Apart from the distribution of the within-day sentiments between 2011 and 2021, the distribution of the periods before and after DST follows a typical pattern and this pattern reflects the within-day sentiment changes of the Twitter users.

One of the main aims of our research is to detect the differences between the periods before and after DST and to discover whether these effects are caused by DST policy. By comparing the two graphs on the Fig 4.2, it is possible to observe within-day

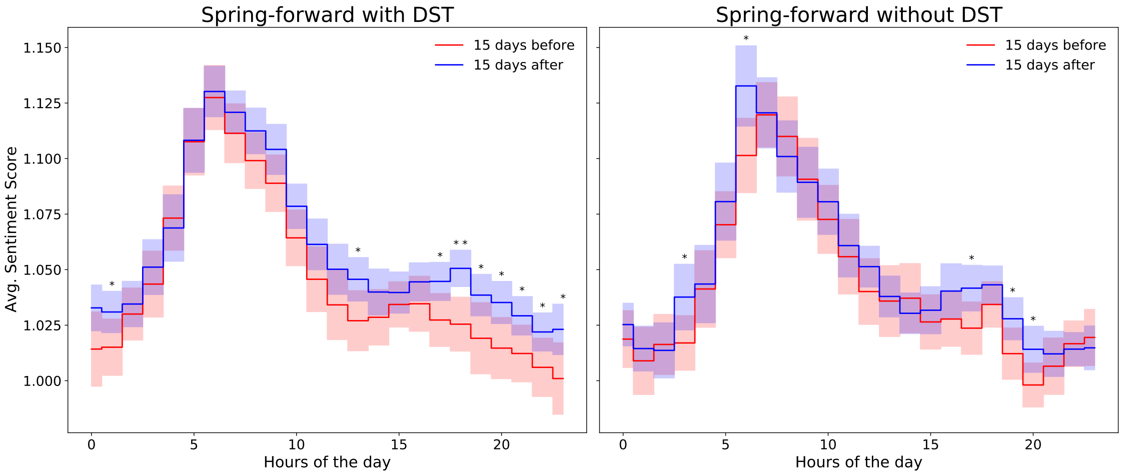

Figure 4.3 **Comparison of the within-day sentiments of the years with and without DST.** Within-day sentiment distribution of the 15 periods before and after spring-forward visualized in the same graph. Stars represents significant difference between two-groups on that hour.

sentiment changes after the DST dates.

Fig 4.3 displays the comparison of the within-day sentiment distributions of the 15 days before and after the DST periods. On the graph on the left-hand side red line represents the pre-DST periods while the blue line represents the post-DST periods. The more transparent red and blue areas around those lines display the 95% confidence intervals of the hourly estimated sentiments. To identify these two groups (15 days before and after spring-forward) numerically, two-sided t-tests were applied each of the 24 hours of the day. Areas with the stars exhibit the hours with a significant difference between the sentiments of the two groups. A single star symbolizes the 95% confidence level while double stars symbolize 99%. According to the results of the t-tests, in the evenings, there is a significant difference between the estimated sentiments of the 15 days periods before and after spring-forward.

This suggests that people tweet more positively in the post-spring-forward periods, especially in the evenings. However, apart from the time shift, these significant differences in the sentiments can be caused by several seasonal factors such as the spring effect and day extension. In order to understand whether these effects are due to DST or other effects, the same periods of the years (after 2016) when the DST policy was not applied were compared. The graph on the right-hand side shows the hourly distribution of the sentiments for the same periods of the years without the DST policy. When the two graphs are compared, it can be observed that the differences in the evening hours before and after DST are more distinct in the graph on the left. As a result, besides seasonal effects, moving the clocks forward also has a positive significant effect on sentiment, particularly in the evening hours.

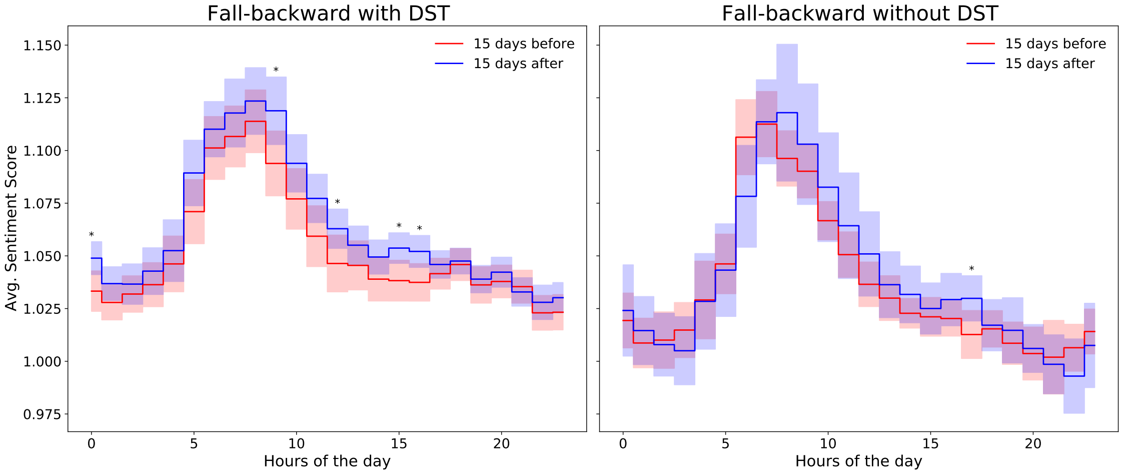

Figure 4.4 **Comparison of the within-day sentiments of the years with and without DST.** Within-day sentiment distribution of the 15 periods before and after fall-backward

Fig 4.4 illustrates the comparison of the within-day sentiment distributions of the 15 days before and after the fall-backward periods. In the years that DST was implemented, during the periods 15 days after fall-backward sentiment scores in the afternoons are significantly higher compared to the periods 15 days before fall-backward. However, this difference remains specific to some hours of the day instead of persisting in a wide time period as in the spring-forward period. The graph on the right-hand side displays the within-day sentiment during the same period of the years without DST policy. The hourly local differences in the years with DST have been absorbed as this graph shows.

4.2.2 How does social clock affect within-day sentiments ?

The prior research question proved interventions like the DST policy have a significant effect on the sentiment of the people. The social clock is affected by daily life events such as the beginning of school, commute time, and end of the shift. The time of those events during the day is directly related to the DST policy. By normalizing the DST effects, it is possible to witness whether the amount of daylight or social time is more influential on people's sentiments.

During the normalization of the clock shift due to DST, the clocks are set back 1-hour in the post-DST period in the spring, while the clocks are forward 1-hour post-DST in the fall. This process aims to compare the effects of sunlight on the sentiment and the effects of the social clock on the sentiment.

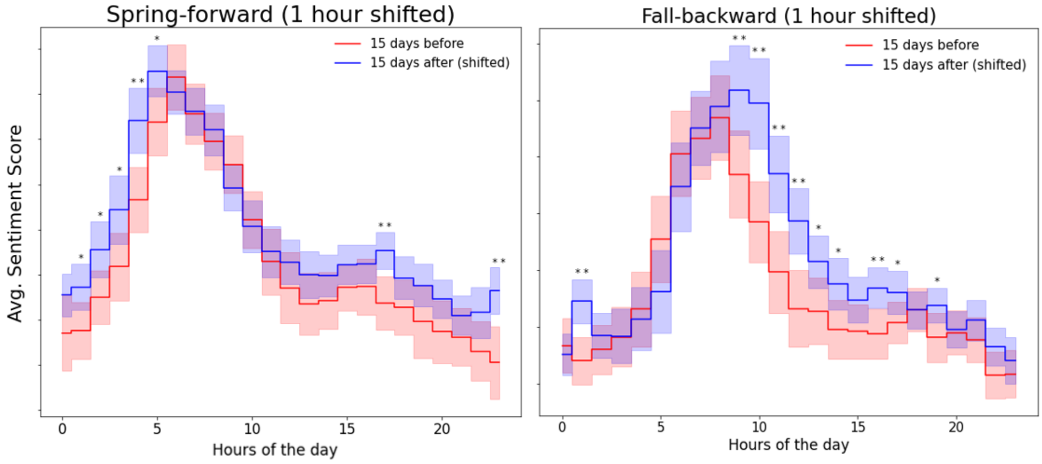

Figure 4.5 **Effect of social clock.** Differences in the within-day sentiment distribution caused by one-hour normalizing the periods after DST

Fig 4.5 compares pre-DST periods with the normalized post-DST periods. Normalizing the post-spring-forward period shown in the left graph, caused a significant difference, especially in the period from midnight to morning. Moreover, there is a local difference in the 5-6 P.M which is the end of the shift for most of the businesses. The right graph proves that normalizing post-fall-backward periods leads to differences over wider periods of the day (from early in the morning to the evening). These significant differences show the hours of the day when people are more influenced by the social clock than by the amount of daylight. The fact that this normalization, performed by taking the reverse 1 hour of post-DST periods, leads to significant differences, is proof that the social clock determined by the policies has a significant effect on our sentiment, as well as our biological clock.

4.2.3 How does DST affect seasonal sentiments?

Within-day sentiment distributions of different seasons vary from each other. These variations can be caused by sunrise and sundown times or different seasonal effects. By visualizing these seasonal distributions it is possible to observe within-day sentiments during the seasons with clock shifts (spring and fall) due to the policy of DST.

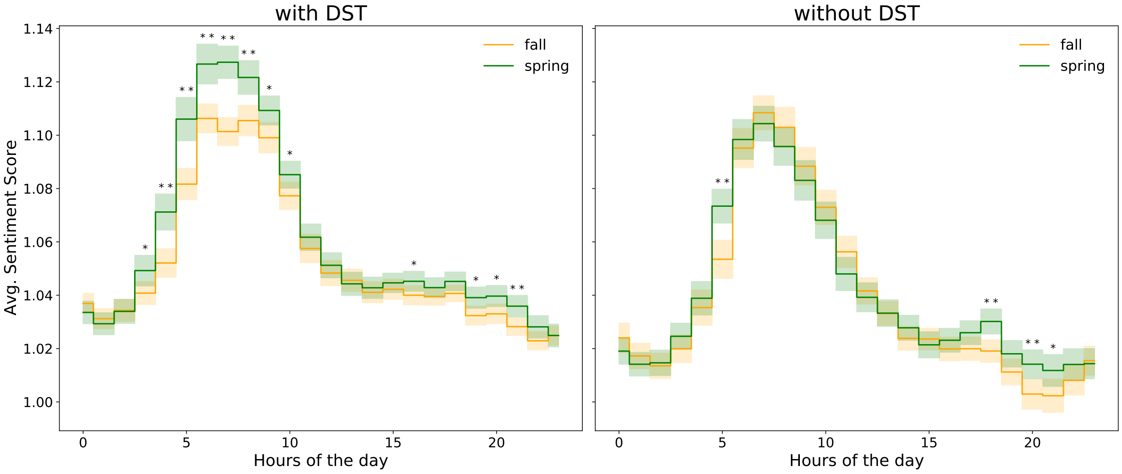

Figure 4.6 **Within-day sentiment over spring and fall.** Distribution of the within-day sentiments during spring and fall months for the years with and without DST

Fig 4.6 compares the within-day hourly distributions of the spring and fall sentiments. Considering the years in which the DST policy was implemented sentiments toward spring mornings are significantly higher than the sentiment toward fall mornings. It is possible to observe the positive differences in favor of spring at sundown hours. The graph on the right compares the same seasons to years in which DST is not applied. Although the significant differences in the evening hours between spring and fall persist, the differences in the morning hours have faded.

Beyond spring and fall within-day sentiment comparisons other seasonal comparisons were performed for both DST and without DST years. Those comparisons indicate sunrise and sunset times that change throughout the year, causing significant differences between hours for the different seasons of the year. Fig A.2 displays later sundown times in the causes a positive significant difference in favor of summer months.

4.2.4 Does societal sentiment reflect important external events ?

During the detection of the major external events both the rapid declines in the time-series sentiment graphs and the most popular 10,000, daily words were used. The duration for the sentiment graphics to be relapsed or the popularity of the event to continue varies from event to event.

Fig 4.7 shows the change in average sentiment score over May 2014. The sharp

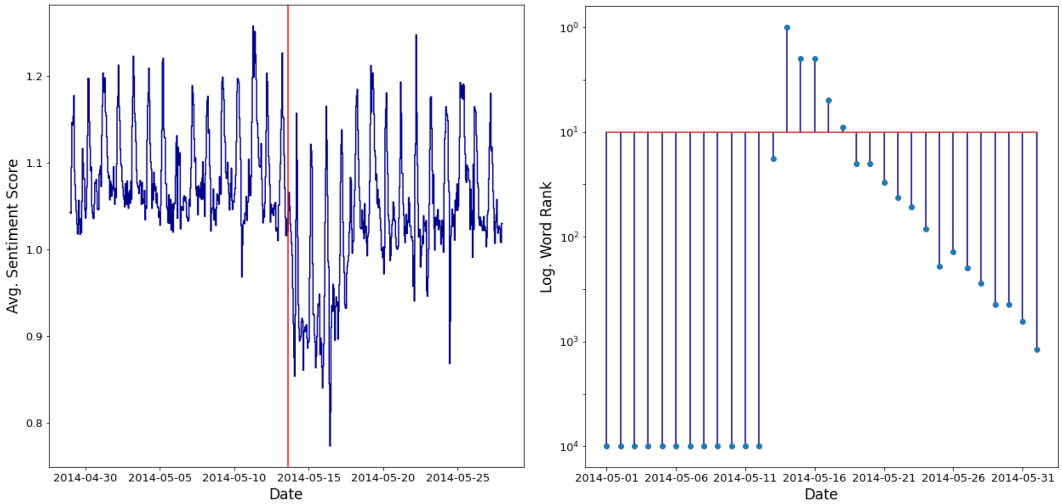

Figure 4.7 **Soma Mine Accident Period.** Sentiment change during the Soma mining disaster on May 15, 2014 (left) Popularity of word Soma among top 10.000 words of the day (right)

decrease started in the second week of May shows the effect of the Soma Mine Explosion that occurred on the 13'th of May. This accident that killed 301 people sharply reduced the average sentiment and this effect faded out after one week. This tragedy also stayed in the Twitter trends for a long time. Fig 4.7 displays the change in the popularity of the word "Soma" over May 2014. Until the day of the disaster, the word 'Soma' is not among the top 10,000 frequently used words of the day. During the period following the event, it stayed above the red line representing the top 10 words of the day for five days.

Fig 4.8 shows the change in average sentiment score over the period between December 2016 and January 2017. The vertical red line represents the new year's day of this year. Before the red line, there is an increase most probably caused by the greeting tweets for the new year. On the new-year eve of this year, a terrorist attack targeted a nightclub in Istanbul named Reina. The sharp decrease started after the red line shows the effect of this event on the average sentiment score. The word 'Reina' stayed among the top 10 most popular words in less time compared to the word 'Soma'. The first one of these days is the new year the incident happened and the second one is the day when the attacker got caught by the police forces.

It is possible to observe the effects of annually recurring events as well as one-off social events on sentiment graphs and the daily most commonly used words. Fig 4.9 shows the sharp rise in the average sentiment score caused by the first day of

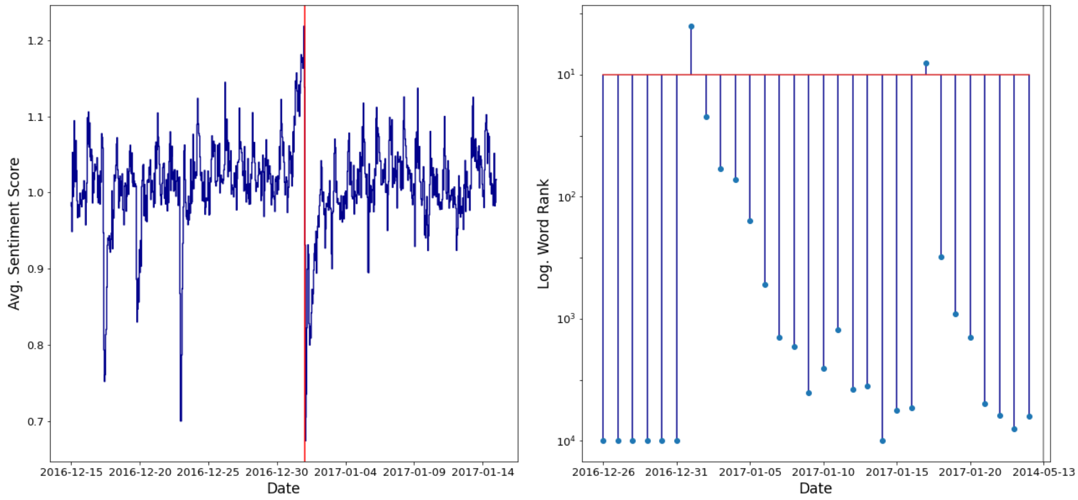

Figure 4.8 **Reina Attack Period.** Sentiment change during the terrorist attack on Reina on Jan 1, 2017 (left) Popularity of word Reina among top 10.000 words of the day (right)

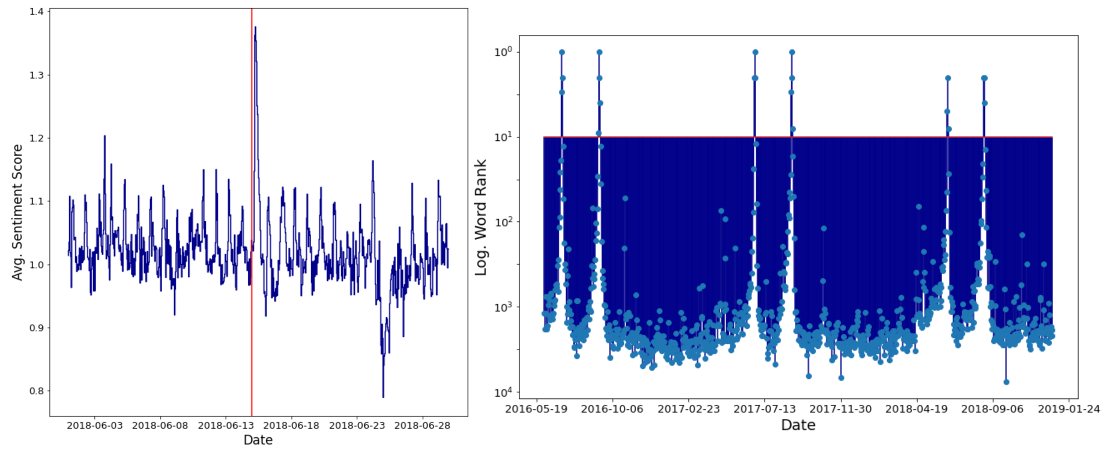

Figure 4.9 **Ramadan Period.** Sentiment change during the Ramadan Feast on 2018 (left) Popularity of word Bayram among top 10.000 words of the day (right)

Ramadan Feast. The graph on the right displays the 3-year change in the ranking of the word “bayram”, which means feast. Word “bayram” becomes one of the top 10 most popular words twice a year. These popularities stems from the annual Ramadan and eid al-adha religional feasts.

5. Conclusion

Within the scope of this study, it is aimed to make inferences on the effects of the Daylight Saving Time (DST) policy, which is one of the most controversial policies, on human psychology by using comprehensive social media data. In order to classify Turkish tweets in the dataset, we proposed a Turkish BERT model, fine-tuned for the sentiment analysis task. Hourly sentiment estimates were obtained, by grouping the scores from the model, and used to analyze the time-dependent changes in the sentiment and the effects of DST on Twitter sentiment.

5.1 Limitations and Development of Research Questions

Aside from the advantages of using social media data to increase the number of observations, it brings its own challenges. By fine-tuning a model with labeled Turkish Tweets from the main dataset we attempted to overcome the difficulties of using noisy Twitter data. Considering that we can detect external events that have taken place in Turkey in the last 10 years, the model we proposed in this study reflects the behaviors of the people. However, there are still some gray areas in our dataset. Cases such as sarcastic tweets, dark humor, or texts from the news that did not include any personal opinions can constitute some ambiguities. When choosing a training set increasing the sample size of those challenging cases can increase the performance of the model.

The most compelling factor encountered when examining the time-dependent change in Twitter sentiment was the effects of external public events. To minimize the effects of these events on the DST policy analysis, daily most popular words on Twitter were collected. However, the removal of the Tweets with the most popular words did not cause significant changes during the external event periods. Until this point of the research, our main focus was daily and weekly average sentiment scores that focused on DST periods. Nonetheless, with the discontinuation of the extraction of external events, hypotheses and research questions centered on within-day sentiment distributions during DST periods.

5.2 Summary of Findings

According to our findings, in a country like Turkey with a busy political agenda, the sentiment of social media users is influenced by major public events. Events such as elections, accidents, terrorist attacks, holidays, and sports events can cause a rapid decline or rapid rise in sentiment score. Moreover, the duration of this effect of those events on sentiment score time-series graphs varies from event to event.

To minimize the effects of these events in DST analyses, in addition to the time-series graphs showing the time-dependent changes of the average sentiment, graphs showing the hourly distribution of the sentiment estimates were also used. When the hourly distributions of 15-day periods before and after DST were compared, moving the clocks forward by one hour has a significant positive effect on people's sentiment in the evening. When this image is created for the years when DST is not applied (after 2016), it is observed that the difference in favor of the post-spring-forward period in the evenings faded out.

As a continuation of within-day sentiment analysis, the effect of the social clock on people's sentiment was investigated by normalizing 15-day periods after DST. As a result of this procedure, it was concluded that the social clock caused a significant difference in the sentiment both in the spring-forward and fall-backward periods of the year.

As part of the analysis that examines within-day sentiment distributions according to seasons, it was founded that during the springs of the years when DST is applied sentiments of the Turkish tweets are significantly higher compared to falls of the same years, especially in the mornings and evenings. When the years in which DST was not applied are examined, it is possible to observe that this difference in favor of spring resumes in the evenings but expires in the mornings.

5.3 Policy Implications

The target of this project is to develop data-driven approaches to policies particularly affecting public health and psychology. As a result of these analyzes, although the negative effects of exiting the DST policy in 2016 cannot be determined individually, it is possible to observe that this withdrawal ended the positive effect of spring-forward at different times of the day. Moreover, the positive seasonal differences in favor of the spring mornings compared to fall mornings also disappeared in the years without policy. These two deductions indicate, that exiting the policy did not cause any positive effect on sentiment and also eliminates the positive effects of the policy. Considering the results of the project, there is strong evidence about returning to the Daylight Saving Time policy will have positive effects on people's sentiments.

BIBLIOGRAPHY

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, (pp. 30–38)., USA. Association for Computational Linguistics.

boyd, d. m. & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230.

Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa-Anke, L., Liu, F., Martínez-Cámara, E., Medina, G., Buhrmann, T., Neves, L., & Barbieri, F. (2022). Tweetnlp: Cutting-edge natural language processing for social media.

Chi, S., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune bert for text classification?

Choudhary, N., Singh, R., Bindlish, I., & Shrivastava, M. (2018). Contrastive learning of emoji-based representations for resource-poor languages. *CoRR*, *abs/1804.01855*.

Choy, M., Cheong, M. L. F., Laik, M. N., & Shung, K. P. (2011). A sentiment analysis of singapore presidential election 2011 using twitter data with census correction.

Devlin, C. L. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, I. A., Scheffer, M., & Bollen, J. (2018). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour*, *3*(1), 92–100.

Golder, S. A. & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878–1881.

Hansen BT, Sønderskov KM, H. I. D. P. S. (2017). Daylight savings time transitions and the incidence rate of unipolar depressive episodes. epidemiology.

Havranek, T., Herman, D., & Irsova, Z. (2018). Does daylight saving save electricity? a meta-analysis. *Energy Journal*, *39*, 63–86.

Jahan, M. S. & Oussalah, M. (2021). A systematic review of hate speech automatic detection using natural language processing.

Karasu, S. (2010). Yaz saatİ uygulamasinin bİnalarda aydinlatma İÇİn kullanilan elektrİk tÜketİmİne etkİlerİ.

Kolchyna, O., Souza, T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination.

Leypunskiy, E., Kıcıman, E., Shah, M., Walch, O. J., Rzhetsky, A., Dinner, A. R., & Rust, M. J. (2018). Geographically resolved rhythms in twitter use reveal social pressures on daily activity patterns. *Current biology : CB*, *28*(23), 3763—3775.e5.

Linnell, K., Alshaabi, T., McAndrew, T., Lim, J., Dodds, P. S., & Danforth, C. M. (2020). The sleep loss insult of spring daylight savings in the us is absorbed by twitter users within 48 hours.

Martín-Olalla, J. M. (2019). The long term impact of daylight saving time regulations in daily life at several circles of latitude. *Scientific Reports*, *9*(1).

Ostic, D., Qalati, S. A., Barbosa, B., Shah, S. M. M., Vela, E. G., Herzallah, A. M., & Liu, F. (2021). Effects of social media use on psychological well-being: A mediated model. *Frontiers in Psychology*, *12*.

Ozdemir, A. & Yeniterzi, R. (2020). SU-NLP at SemEval-2020 task 12: Offensive language IdentifiCation in Turkish tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, (pp. 2171–2176)., Barcelona (online). International Committee for Computational Linguistics.

Pellert, M., Metzler, H., Matzenberger, M., & García, D. (2021). Validating daily social media macroscopes of emotions. *CoRR*, *abs/2108.07646*.

Raheman, A., Kolonin, A., Fridkins, I., Ansari, I., & Vishwas, M. (2022). Social media sentiment analysis for cryptocurrency market prediction.

Rishi, M. A., Ahmed, O., Perez, J. H. B., Berneking, M., Dombrowsky, J., Flynn-Evans, E. E., Santiago, V., Sullivan, S. S., Upender, R., Yuen, K., Abbasi-Feinberg, F., Aurora, R. N., Carden, K. A., Kirsch, D. B., Kristo, D. A., Malhotra, R. K., Martin, J. L., Olson, E. J., Ramar, K., Rosen, C. L., Rowley, J. A., Shelgikar, A. V., & Gurubhagavatula, I. (2020). Daylight saving time: an american academy of sleep medicine position statement. *Journal of Clinical Sleep Medicine*, *16*(10), 1781–1784.

Roenneberg, T., Wirz-Justice, A., Skene, D. J., Ancoli-Israel, S., Wright, K. P., Dijk, D.-J., Zee, P., Gorman, M. R., Winnebeck, E. C., & Klerman, E. B. (2019). Why should we abolish daylight saving time? *Journal of Biological Rhythms*, *34*(3), 227–230. PMID: 31170882.

Safaya, A., Abdullatif, M., & Yuret, D. (2020). KUISAIL at semeval-2020 task 12: BERT-CNN for offensive speech identification in social media. *CoRR*, *abs/2007.13184*.

Sayce, D. (2020). The number of tweets per day in 2020.

Schweter, S. (2020). Berturk - bert models for turkish.

Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. (2019). Attentional encoder network for targeted sentiment classification.

Statista (2021). Countries with the most twitter users 2021.

Tong, L., Zhang, X., Zhang, Q., Sadka, A. H., Li, L., & Zhou, H. (2019). Inverse boosting pruning trees for depression detection on twitter. *CoRR*, *abs/1906.00398*.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *CoRR*, *abs/1403.7400*.

Varol, O., Ferrara, E., Menczer, F., & Flammini, A. (2017). Early detection of promoted campaigns on social media. *CoRR*, *abs/1703.07518*.

Xin Tong, Yixuan Li, J. L. R. B. L. Z. (2022). What are people talking about in backlivesmatter and stopasianhate? exploring and categorizing twitter topics emerging in online social movements through the latent dirichlet allocation model.

Yang, X., Chen, B.-C., Maity, M., & Ferrara, E. (2016). Social politics: Agenda setting and political communication on social media. In *Lecture Notes in Computer Science*, (pp. 330–344). Springer International Publishing.

Yeniterzi, R. & Oflazer, K. (2010). Syntax-to-morphology mapping in factored phrase-based statistical machine translation from english to turkish. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 454–464.

APPENDIX A

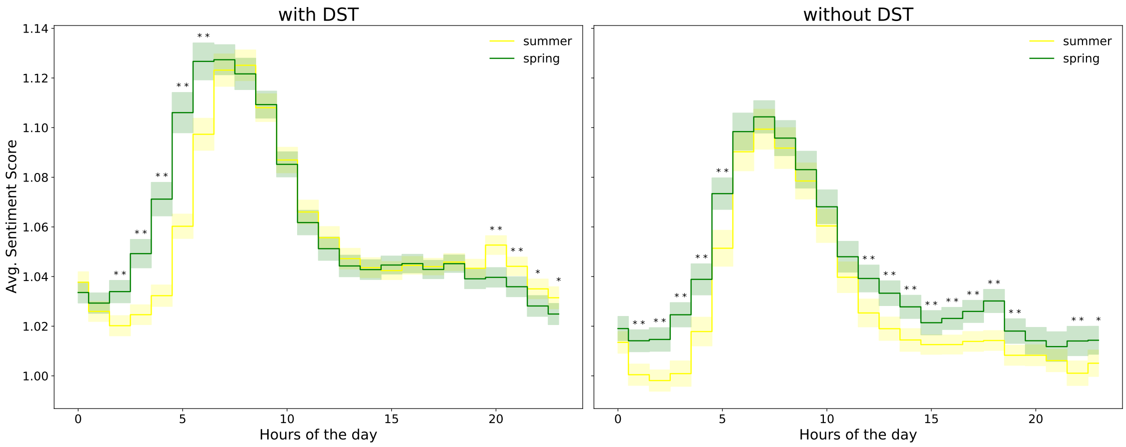

Figure A.1 Comparison of the spring and summer within-day sentiments of the years with DST and without DST

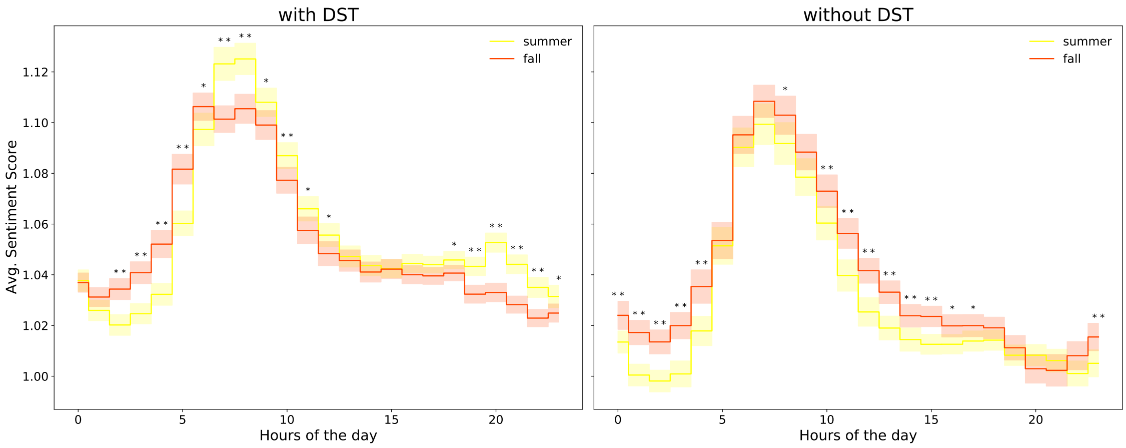

Figure A.2 Comparison of the fall and summer within-day sentiments of the years with DST and without DST

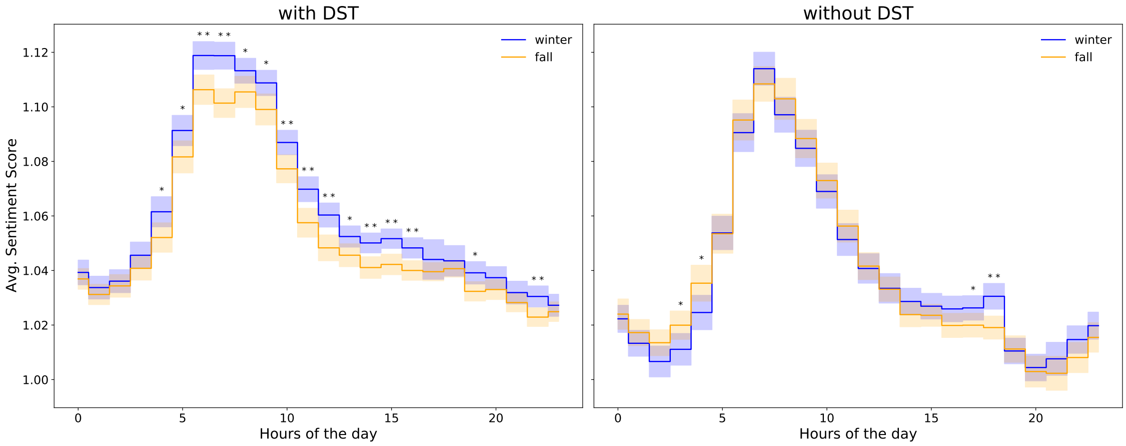

Figure A.3 Comparison of the winter and fall within-day sentiments of the years with DST and without DST

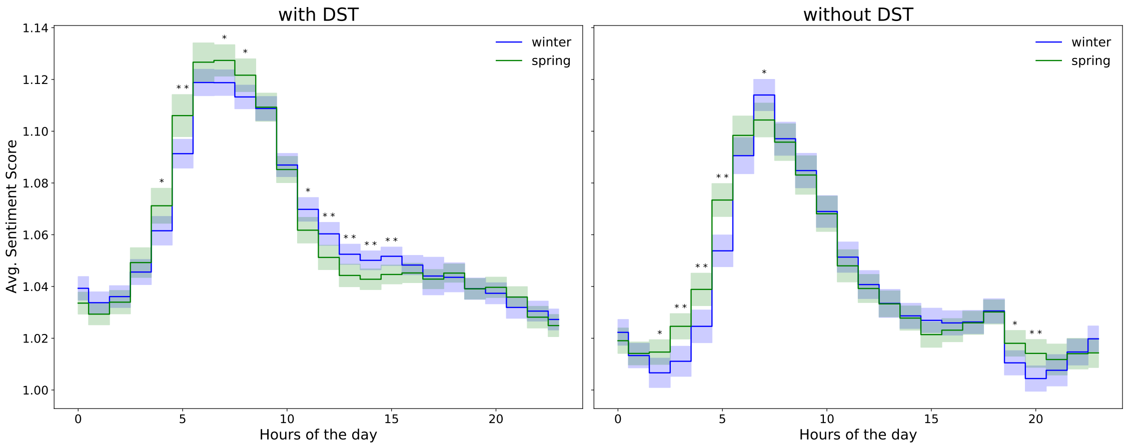

Figure A.4 Comparison of the spring and winter within-day sentiments of the years with DST and without DST

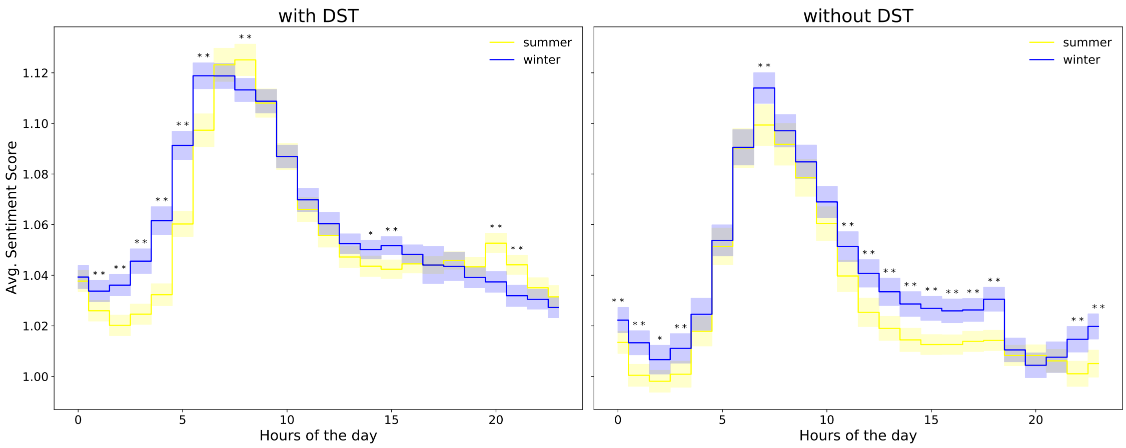

Figure A.5 Comparison of the winter and summer within-day sentiments of the years with DST and without DST