THE IMPACT OF TELEVISION COMMERCIALS AT SPECIFIC DAY PARTS AT SPORTS EVENTS ON WEBSITE VISITORS

PREDICTING SHORT-TERM WEBSITE TRAFFIC

BRAM BEIJER

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

Word count : 8733

STUDENT NUMBER

2085392

COMMITTEE

prof dr. M.Louwerse
prof. dr. G. Chrupala

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

January 13, 2023

ACKNOWLEDGMENTS

I would like to express my gratitude to my primary supervisor, Max Louwerse, who guided me throughout this project. I would also like to give a special thanks to my company supervisor, Charissa Bos, for her knowledge and the time she took for guiding me throughout the thesis. This thesis is especially dedicated to my family and girlfriend, I know they are proud.

DATA SOURCE/CODE/ETHICS STATEMENT

Work on this thesis did not involve collecting data from human participants or animals. The original owner of the data and code used in this thesis retains ownership of the data and code during and after the completion of this thesis. The author of this thesis acknowledges that they do not have any legal claim to this data. Data supporting this thesis are available from Springbok Agency.

CONTENTS

THE IMPACT OF TELEVISION COMMERCIALS AT SPECIFIC DAY PARTS AT SPORTS EVENTS ON WEBSITE VISITORS

PREDICTING SHORT-TERM WEBSITE TRAFFIC

BRAM BEIJER

1 ABSTRACT

Over the last years, the number of commercials during sports events have grown in comparison to traditional marketing. Due to the increase of these commercials, knowing what the return on investment (ROI) at a specific time is becomes more important. This also means targeting marketing becomes more crucial. An increase of website traffic is a form of ROI. Therefore, understanding website traffic is helpful in specifying target marketing. This study will highlight the extent to which time series models ARIMA, LSTM, and XGBoost can predict website visitors the best during a Formula One sports event with specific commercials. The first dataset that was used is generated from a Google API that recognizes logos. The other set was made available by a supermarket that collected website traffic. To start, the data is described, cleaned and the number of website visitors is divided in different timestamp categories. To measure the difference in short-term behavior. Furthermore, some more improvements, like making sure that the data is stationary, were made to get the data ready for the three forecast time series models. Evaluation metrics were used to compare the error rates and the predictive power of the models per different timestamp. The best short-term timestamp is 1 hour after seeing the logos and is best predicted with the ARIMA model. However, it cannot be said that the morning, afternoon or evening influence the number of website visitors during a Formula One event. On the other hand, there are still some critical points that require additional research. Certain outliers are included that might have a big impact on results. Extra marketing elements that contribute to the impact of website traffic can strengthen the research, since it is difficult to conclude based on one marketing element, sponsorship.

2 PROJECT DEFINITION, MOTIVATION & RELEVANCE

Sponsorships are rapidly progressing in comparison to traditional marketing in the sports category. It becomes harder to get the viewer's attention with normal traditional television advertisements, radio and television spots (Grohs, 2016). Furthermore, during COVID-19 almost three-quarters of the brands still extended their sponsorships (Cobbs, Jensen, & Tyler, n.d.). This applies, among others, to the supermarket branch, that are investing in this type of marketing as well.

The supermarket branch has had questions for the last three years about the television events sponsorship budget and impact, and if it might be possible to evaluate the results within the marketing performance (Springbok, 2016). In this case, the television commercials at a Formula One event may influence market performance. Currently, supermarkets often only know the costs but do not know other numbers, like website visitors, to quantify sponsorships. However, in certain cases, the supermarkets will get the ratings from television events to calculate the Return on Investment (ROI). ROI is calculated by dividing the costs by the number of viewers (Jensen & Cobbs, 2014). To be more specific and informative than just getting the ROI, logo counting, or sponsor recall in return, more quantifiers might be needed to get more insight in the sponsorships (Cameron, n.d.). An example of an extra insight is the understanding of website traffic and acting on the change of website visitors created by the commercials. Also due to the growth in online marketing, this study is set up to calculate the impact of commercials on website visitors and to eventually create value for the supermarket branch. At first, in this research proposal classification models were described as the algorithms that provided the best results for this topic. Based on the data generated from the Google API, classification was not the right option to predict website visitors. New techniques were researched to solve this problem.

Understanding online website traffic can be modeled mathematically using time series models. Time Series Forecasting (TSF) is an important application area covering many different fields, from weather forecasting to indicators of economic trends to demand-driven power plant construction and therefore also website traffic on specific times. TSF became a strong research precedent and has attracted the attention of several scientists around the world (Zhou, Wang, Huang, & Liu, 2021), (Makridakis, Spiliotis, & Assimakopoulos, 2020). Not only are there many scientific works on this topic, but there are also many competitors. These activities have led to the development of several real-time sequence forecasting methods.

Furthermore, in 2021 Casado-Vara, Martin del Rey, Pérez-Palau, de-la Fuente-Valentín, and Corchado (2021) also tried to forecast online web traffic with the time series model, long short-term memory (LSTM). In that research, data is generated by scraping web pages. The results fitted well, but predicting anomalous behavior was hard. What might occur due to the use of only one model. Cameron (n.d.) predicted the length of commercials on customer behavior by using a survey and among others, eXtreme Gradient Boosting (XGBoost) model, what generated low results on the short-term behavior. With other types of data, predicting continuous quantity instead of classes and observing website visitors might change the results. This will be done for the Formula One sports events with different models, to increase the quality of the results.

To fill in the gap of researcher Cameron (n.d.), a representative sample is created in this research, since a supermarket is delivering real-time data of their collected website results. With the Google video intelligence API, logos of supermarkets and other related logos to the supermarket can be detected from television events. After combining these data sources, LSTM, XGBoost, and Auto-Regressive Integrated Moving Average (ARIMA), time series forecasting machine learning models predict results that will be important for target marketing and fill in the gap of (Casado-Vara et al., 2021).

Finally, the most important purpose of the supermarket branch is to gather more insights into the results of sponsorships and an increase of potential customers. The relevancy will increase if more specific data, results in stronger outcomes. Worldwide usage of this research is possible for companies that also sponsor during television sport events, gather data, and have a website to increase their performances. Or new potential business opportunities can be created that predict the value of sponsor deals during sports events. Also, more research can be done into the effect on other quantifiers (sales), to optimize a company even more.

Based on the outlined relevance and research gaps the following main questions with sub-questions have been created: *To what extent can the number of website visitors be predicted, based on television commercials during sports events?*

RQ1 *which of the researched time series machine learning models can predict the number of website visitors most accurately?*

The researched time series machine learning models in this study are ARIMA, LSTM and XGBoost. These three will be compared, based on different type of evaluation metrics.

RQ2 *Which of the researched time categories can be used to predict the website behavior best?*

Time categories are consisting of the number of website visitors and divided based on types of short-term behavior. These short-term categories are visitors before and after seeing commercials during a Formula One race. The following categories of short-term behavior are used: 1 minute before seeing the logo and 1 minute, 10 minutes and 1 hour after seeing the logo.

RQ3 *To what extent do day parts influence the number of website visitors prediction when the sports event is happening?*

Question 3 will give more insight into the influence of day parts on the number of website visitors. The different day parts are the morning, afternoon and evening. The morning starts at 6:00 a.m. and ends at noon. The afternoon starts at 12:01 a.m. and ends at 6:00 p.m. Lastly, the evening starts at 6:01 p.m. and ends at midnight. No Formula One sports event take place at night, so the night is left out.

The main finding of this research is that the model ARIMA can fit and predict the data best. Mainly, it fits the data of the timestamp categories '1 hour after seeing the logos' the most. The best time category also applies to the LSTM, but not for XGBoost. Despite these results, there is a weak correlation between the growth of website visitors and the day parts. These results will answer the questions above.

3 LITERATURE REVIEW

The following sections will discuss the current state of knowledge on the influence of marketing and sales. Also, other papers will be discussed that tried to predict customer behavior based on television events or commercials. Besides that, the used machine learning models will be described and compared.

3.1 *Confounding influences customer behavior*

First of all, forms of marketing such as campaigns, discounts, and television commercials influence customer behavior, and thereby they influence the website traffic as well. Television commercials have a major impact on awareness (AWR), interest (INT), and conviction (CON) stages of customer behavior. In case of short-term behavior, it is important to often give reminders of the brand to consumers (Sama, 2019). So, in short, the minutes straight after seeing a logo on a sports event and right before the commercial break are the most important minutes, to avoid the influence of commercial breaks. So, only these minutes in between will be used to tackle this influence. To see the difference, these short-term time categories will be compared to a longer-term category. Furthermore, the weekends with specific discounts and campaigns will be excluded. Specific products that will have a connection with sports events affect the sales of supermarkets when discounted. Additionally, campaigns that are for example, dedicated to a celebrity that is linked to the supermarket, will increase the number of website visitors. An example, is the Formula One racer Max Verstappen from Red Bull racing, who is connected to the supermarket. This means that the Red Bull logo will also be added to the data besides the supermarket logo (Voorveld, Araujo, Bernritter, Rietberg, & Vliegenthart, 2018).

Koronios, Ntasis, Dimitropoulos, and Ratten (2021) researched specific aspects that influence behavior related to sponsorships during sports events. In this study the researchers looked at several elements, who can influence the behavior due to the commercials. For example, purchase or attitude behavior towards sponsors. They delved deeper into the positive effects, rather than just looking at ROI. It demonstrates that actual purchase behavior for sponsors' goods and services is significantly influenced by a variety of antecedents as opposed to just purchase intentions. Their research demonstrates that sponsors must fully participate in the sponsorship process.

3.2 *Online traffic*

Second, in the past decade, 20% of the sales made in 2018 for the UK were online sales. Nowadays, retailers are busy with reaching new customers online and improving the customer experience (Dolega, Rowe, & Branagan, 2021). An increase in online sales will often mean that the website traffic positively changes. On the other hand, the path of a customer through the website differ from customer to customer, and it can take long before the purchase takes place. 90% of the people will not buy anything at their first website visit. A lot of decisions are being made in the mind of a website visitor before they actually buy something. Marketing influences this website traffic and changes the behavior and train of thought of a visitor when seeing the logo, product, or a name more often (Dolega et al., 2021), (Kakalejčík, Bucko, & Danko, 2020).

3.2.1 *Main day parts online traffic*

To follow up on this, the parts of the day also influence website traffic. Avraam, Veglis, and Dimoulas (2021) researched the difference in website visits between day parts, weeks or weekends. Zooming in on day parts reveals that the highest peak lays around 10:00 p.m., the start of this peak begins at 6:00 p.m. Besides this, events on television will be broadcast on specific times to reach the right audience, and reach the highest peak of users for the best results.

3.3 *Time series Technique*

An ordered sequence of values recorded at equal intervals over time is called a time series. There are two parts to the analysis of time series. The first step is to discover the underlying pattern and structure of the data. The second part, focuses on how to fit a model to make predictions for the future. Time series analysis is used for a lot of things, like economic forecasting, quality and process control or census analysis. Since more companies are interested in real time monitoring of their data, time series analysis has been extensively used to forecast their results of target marketing. Additionally, this type of monitoring provides previously recorded data that can be utilized for website traffic analysis and forecasting (**?**). Decomposing a time series into the three parts listed below is a common method for doing so. **Trend**: The general trend that the variable shows over the course of the observation period without taking into account seasonality. **Seasonality**: is the variable periodic fluctuation during an analysis. It includes time, magnitude, and direction as well as stable effects. **Residual**:

The remaining part of a time series. If there are a lot of residuals it might influence the trend and seasonality. Fanoodi, Malmir, and Jahantigh (2019) researched the importance of a strong trend and seasonality, which means that the data must be stationary. The high's and the low's will be the same over time in a horizontal line. The prediction model's error will rise if the data pattern is not stationary.

Time series analysis typically falls into two categories: univariate and multivariate. Univariate time series is a single observation by time in sequence, for instance, the number of website visitors every minute. When a group of time series variables and their interactions are to be considered, multivariate time series is used. Univariate time series analysis is the primary focus of this study (Deb, Zhang, Yang, Lee, & Shah, 2017).

3.4 *Predictive models*

To predict website visitors time series models ARIMA, XGBoost and LSTM were used. These three models can predict the number of website visitors on a certain time and display the increase or decrease overtime.

An ARIMA model is often called the most general time series modelling technique, and is therefore frequently used as baseline. The base of an ARIMA model is that it describes the trends and seasonality patterns in time series. It has a great advantage in the simpler (univariate) time series forecast models by identifying the existing linear structure in the data (Khashei & Bijari, 2011), (Shelatkar, Tondale, Yadav, & Ahir, 2020). The ARIMA model has demonstrated superior accuracy and precision in predicting the next steps in time series. However, it will often be outperformed by the LSTM model (Siami-Namini & Namin, 2018).

The XGBoost belongs to the top of machine learning algorithms and is often used by winning teams in the Kaggle competitions (M. Chen et al., 2019), (T. Chen & Guestrin, 2016). In addition, for time series prediction XGBoost perform extremely well, because of his memory resources usage and efficient computing time (Abbasi et al., 2019). XGBoost is a strong implementation of gradient-boosted decision trees. The main reason that XGBoost is effective, is that it combines a number of relatively weak decision trees to obtain a more powerful prediction (Nielsen, 2016). The order in which the feature interactions occur is determined by the depth of the decision tree (An & Ren, 2020). Conversely, under-performs in cases with big data and very high feature dimensions (Yang & Zhai, 2022).

The main idea of the LSTM model architecture is the memory cell, which can maintain the data for a long period of time and regulates the data flow into and out of the cell. It is a better developed time series model than the ARIMA model (Li & Zhang, 2018). It also leads to better performances of time series than the ARIMA model most of the time. Because of its performance for time series and its performance against other machine learning algorithms (Siami-Namini & Namin, 2018), the LSTM algorithm is chosen for the task researched in this thesis.

3.5 *Predictive modeling in commercials*

The effect of commercials during sport events (sponsorships) is an under-researched problem. There have been studies that tried to predict customer behavior or were searching for a correlation between customer behavior and commercials. And where deeper insight can expose more influencing factors (marketing). For example, if the success of performance in sports influences the return on investment (Cobbs et al., n.d.). Furthermore, Carreón, Nonaka, Hentona, and Yamashiro (2019) showed with machine learning models which also included the XGBoost that short-time purchase behavior generated low predictions result based on television commercials. The models and method that will be used differ from those of this study, to see different results when the concept of the research is similar. A few other researchers such as Casado-Vara et al. (2021) researched web traffic for web server providers to have a better control of their web pages using forecast time series. In this study, LSTM is compared with Recurrent Neural Networks. Concluded that the LSTM performed better than an RNN. With the LSTM, the study generated a great accuracy for large and small data sets, despite the fact that the data had bad patterns. In addition, Shelatkar et al. (2020) researched the possible manners to forecast web traffic with different time series models. ARIMA, LSTM and RNN were compared, which resulted in the fact that LSTM could capture the patterns best. Lastly, the XGBoost has also been compared to the LSTM, as the LSTM often performs better than other time series models. In this study Luo, Zhang, Fu, and Rao (2021) COVID-19 transmission is predicted with different time series. This study shows that the LSTM performs better with a error rate percentage of 2.32%, unlike the XGBoost, which has a error rate percentage of 7.21%. Although the comparison between LSTM, ARIMA and XGBoost is not common on this topic. Often, either only ARIMA is compared with LSTM or LSTM with XGBoost.

4 METHODOLOGY & EXPERIMENTAL SETUP

This section starts with describing the data sets that were used. The preprocessing, like tracing seasonal trends and outliers that were performed for the analysis phase. Next, the supervised learning problem and the data scaling is explained. Afterwards, the data will be split, the hyperparameters are being tuned and the results will be compared with the evaluation metrics. Lastly, the used package will be described.

4.1 *Algorithms*

4.1.1 *ARIMA*

Auto Regressive Integrated Moving Average (ARIMA) includes the moving-average (MA), and auto regression (AR) models. Both are statistical time series forecast models. MA depends on previous forecast residuals, who are needed for the next forecast. On the other hand, AR is based on lagged values of the data that were used for forecasting. These formulas belong to the models:

MA (1), where the Yt is predicted by the lagged values of the error ϵ_t, θ is the input of the autocorrelation from the errors, and q stands for the number of lags.

$$Y_t = c + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}|(1) \tag{1}$$

AR (2), where the Yt is predicted by the lagged values of Yt, c stands for the constant, ϕ is seen as the magnitude of the autocorrelation, p is the quantity of lags, and ϵ_t is the error.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t|(2) \tag{2}$$

These models will give the best results when the data is stationary. This means that there is no trend or seasonality in the data and for non-stationary vise versa. With the Augmented Dickey-Fuller Test it is possible to test if the mean and variance are consistent over time and whether they are stationary. If the test shows that the data is not stationary, the following can be done: subtracting the previous observation from the current observation. This method is called Differencing (3) (Schaffer, Dobbins, & Pearson, 2021). The first observation(t) is the starting period, and the observation(t-1), means that it moves every step further away from the starting period. This have to be done over and over again to extract the trend out the data.

$$difference(t) = observation(t) - observation(t-1)|(3)$$

Besides this, the parameters of the ARIMA are p, d and q. P stands for quantity Autoregressive values (AR). It is a regression model that uses the correlation between an observation and a number of lagged observations (an observations that will occur later in the time series). The amount of significant partially correlated lags is the best value for p. Q is the Moving Average (MA), to find the best value for q, subtract the mean of the first lag and multiplying this cost with the preceding lag minus the mean. This technique is repeated for all k lags and is summed, forming the numerator of the formula. The denominator is the variance of the authentic series, ergo: the sum of the squared distinction among the mean and every lag. D is the Differencing order, it indicates the amount of times required to bring a sequence to a sort of constant, that remains the same over a period. If the data is stationary, d should be 0. The non-seasonal part of the model are the parameters (Benvenuto, Giovanetti, Vassallo, Angeletti, & Ciccozzi, 2020). P indicates the connection order of the time series and its past, q is the connection of the series to the active factors in its design. The time series analysis is designed in more than one stage. It starts with finding the best values of the parameters with the Autocorrelation Function (ACF) and Partial Autocorrelation function (PACF). These plots of the functions will give a general display of the time series, its properties, and trends. This display is often the basis of selecting the best model. The next step usually checks if p and q (for autoregressive and moving average value) can remain in the model or must be removed. Lastly, it checks whether the residuals are random and normal distributed (Shamsnia, Shahidi, Liaghat, Sarraf, & Vahdat, 2011).

4.1.2 *LSTM*

Recurrent Neural Network (RNN) has a special distinctive feature, in comparison to another Neural Network. RNN has hidden units that store information for certain periods of time, which can influence later outputs. A special memory, where the data flows in sequential trough the time steps (Sak, Senior, & Beaufays, 2014). However, when processing backpropagation through the algorithm, the gradients of the neurons may become too large ('exploding') higher than 1 or too small ('vanishing') lower than 1 (Alzubaidi et al., 2021). Long Short-Term Memory (LSTM) is a form of RNN but can deal with this issue by bridging time delays. LSTM replaces the hidden units by cells with extra properties, an extra state, and certain gates. To control every cell separately from each other and move in between these cells. These cell states consist out of three states: the forget gate, looks at the previous cell state, and makes the decision to dispose it because it is not relevant anymore or the gate can keep it in. It will get a 0 when disposing the cell and a 1 when remembering the gate following the sigmoid function. The equation of this gate is formulated in equation (3), where ft the output is between 0 and 1. $t-1$ is linked to the previous cell, xt is the current cell, ht stands for the hidden unit, b is the bias, and the rest (W, U) are weights. The σ stands for the sigmoid function (Farzad, Mashayekhi, & Hassanpour, 2019). This is the same for all gates except the input values (f, i, o).

$$ft = \sigma(W(f)xt + U(f)ht - 1 + b(f)) \tag{3}$$

The input gate determines if the current cell has enough important information to keep it or not. These potential values to see the importance are calculated with the following equation (4). Where the calculation is constructed in the same way.

$$it = \sigma(W(i)xt + U(i)ht - 1 + b(i))) \tag{4}$$

The output gate calculates the next hidden state. In other words, this is the output of the current time step. It is also constructed in the same way as the other layers.

$$ot = \sigma(W(o)xt + U(o)ht - 1 + b(o))| \tag{5}$$

During the process of the forget and input gate the cell state (ct) is formed. This was needed to stand as input value for the input and output gate. Created with this equation (6), where the tanh function overcomes the vanish problem.

$$ct = ftct - 1 + ittanh(W(g)xt + U(g)ht - 1 + b(g)) \tag{6}$$

Lastly, the cell stats duplicates, one will be used forming the hidden unit via the tanh activation, while the other one is used for the next LSTM cell (Smagulova & James, 2019). A display of the process can be found in figure 1. Furthermore, one dense layer located as the last layer to have one single output neuron (Yu et al., 2019).

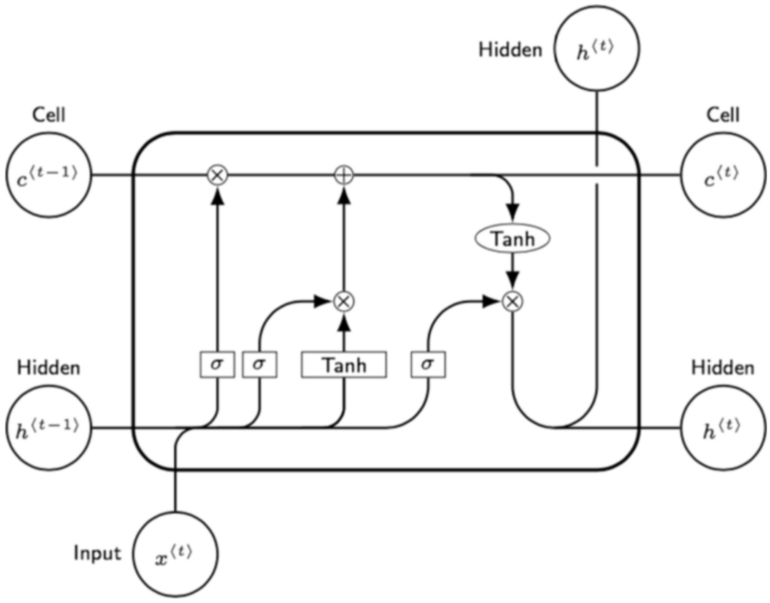

Figure 1: LSTM architecture, the process of the gates and functions of a LSTM model

Sagheer and Kotb (2019)

4.1.3 *XGBoost*

The eXtreme Gradient Boosting (XGBoost) is a gradient tree boosting algorithm. It provides parallel tree boosting. It is a supervised learning model and not made for time-series forecasting. However, this algorithm is often used in competitions, where most of the time, it predicted better than, for example, a Neural Network. XGBoost is based on three main objects constructed in the following manner. By first creating the Regularized Learning Objective, shown in equation (7) (T. Chen & Guestrin, 2016).

$$y\hat{}i = (x_i) = \sum_{k=1}^{k} f_k(x_i), f_k \in F \tag{7}$$

Where $\sum$ the sum is, x_i the input, every f_k stands for an independent tree structure and $\hat{y}$ is the predicted output. $\in$ means that the formula is an element of F. The F stands for the space regression trees. It is about the structure of the trees, the number of leaves, the summed up scores of the leaves, leaf weights and the continuous value of each leaf. A convex loss function calculated the difference between the actual and predicted value. Minimizing this convex loss generated the best results. Afterwards, Gradient Tree Boosting will be created. This is the training part of the model, adding the best f value that comes from the difference between the prediction $\hat{y}$ and target y, penalize the complexity to avoid overfitting by picking a model with simple functions. Lastly, selecting the most optimal weights to structure the best tree. Finally, Shrinkage and Column Subsampling is applied within XGboost to avoid overfitting as well. Shrinkage is similar to a learning rate, that scales the new weights after every tree boosting. Subsampling is selecting a subset (columns) of the data to reduce data. It also increases the computational power (Ogunleye & Wang, 2019).

4.2 *Experimental Setup*

4.2.1 *Dataset description*

The first raw dataset is collected from Google Cloud BigQuery environment and is real customer website traffic data of a supermarket company. The data is collected via Google Tag Manager, which follows the traffic on the website, developed by Springbok Agency (Springbok, 2016). This Big Query table consists of 400+ thousand rows and 17 columns (features). The 400+ thousand rows are website sessions from one day. When a user went to the website, it counts as one session. This is happening almost every second. For clarification, this dataset will be defined as historical data since the dataset will not update in real-time, but day by day. However, the data is collected on the 17th of October. What means that only the data till this date will be analyzed.

The data about the commercials during the sports events is created by a model that recognizes all the logos, which also includes the supermarket in question. This model is based on the Video Intelligence API from Google. It generated a data frame with the logo, entityId and the timestamp of appeared logos in seconds from sixteen races of approximately thirty-six hours of racing. This resulted in 108737 rows with logos that were transferred to BigQuery. The races are from the 2022 season to get the most relevant results. Besides, these races were selected while this research started during the season. This means that a selection of data is made when the API started running.

4.2.2 *Data pre-processing*

Data selection

Of the data coming from the supermarket, all features have been removed in BigQuery except the visitStartTime (in epoch timestamp) and visitId. The visitStartTime will be used for the join between the data sets. Both the data sets have an important feature time. VisitId shows every person (unique id) that visited the website. From the Formula One data every feature is selected.

Only the website visitors during the sport events were necessary. With Big Query, visitors were selected between the start and the end of the races. The selection exists of data from sixteen races, only from Sundays and decreased to 130010 thousand samples. The Tag manager tool collects data of every second based on website traffic. However, when there is no website visitor it will not collect anything and it skips that specific second. Seconds with the value *null* were implemented to also get the logos that passed by in these seconds. To work with these seconds, the value *null* is changed to 0, which means 'there were zero website visitors on that second'.

From the Formula one data only the logos from the corresponding supermarket and Red Bull are selected. The Red Bull logos were also selected because of Max Verstappen, a famous Formula One driver from the Netherlands who drives for Red Bull team and is driving around a car on which the logo Red Bull is displayed in large letters. In addition, Red Bull products are also sold in this particular supermarket. This could mean, based on this source that seeing a logo or person could be associated with a product or supermarket. From the sixteen races 4487 rows were now left over when selecting these logos.

Time transformation & determination

After collecting these target data and related data, the time data of the sports events were converted to epoch timestamps by adding the time the logo appeared to the epoch of the start of the race, to join them together. Only related rows of website traffic that occur between the race times (GMT timestamps) plus 1 hour were selected. This extra hour has been chosen to examine results after the event, when people are often done with enjoying the event and have some time left. To examine if the logo of the supermarket and the logo that is related to the supermarket influence website traffic, the epoch second timestamp is converted to a minute timestamp. The number of occurrences of the supermarket logo and related logo within a minute will be calculated. This was done to better determine whether seeing the logo affects the number of visitors minutes after seeing it. The following categories have been drawn up for this purpose: the number of website visitors after 1, 10 minutes and after 1 hour were added to the dataset. These categories were used to compare short-term impact with a longer-term category impact. This is done to get an insight into whether seeing a logo more often has more impact. The baseline for these categories and the study is the number of website visitors before seeing the logos to notice the difference. This time category is least affected by the logos. To investigate the real impact of the logos, only the minutes where a logo appears were taken into account. The remaining minutes were removed. Lastly, the difference between the baseline and the other categories was calculated. To generate a single feature as the input, for a univariate time series.

Outlier detection

An outlier is often defined as: "observations or patterns that are different from the rest that it is suspected and not normal behavior because most of time made these are made by someone else technique" (Hawkins, 1980). Outliers are also called anomalies in time series. There are also two different sort of anomalies, positional anomalies mentioned refers to observations being anomalous in comparison to the complete data set. A contextual anomaly depends on the context of the data. E.g., a high visitor amount is not anomalous during daytime, but it is during the night. The data is stationary, so only positional anomalies were taken into account. In this study with quite a small dataset these anomalies can be found with simple statistics techniques and graphs. The highest values that differ the most were examined more to see if the difference can be explained by holidays, special weekends, or other campaigns from the supermarket. In figure 2, the amount of visitors per race are being displayed. Three races are remarkable, the race in Monaco on the 29th of May between 2:00 p.m. and 5:15 p.m. GMT, the race in Australia on the 10th of April between 06:00 a.m. and 08:12 a.m., and the race in Miami on the 8th of May between 8:30 p.m. and 10:49 p.m. GMT. Around the time of the race of Monaco, more races were being held during a similar weather season, where no outliers can be found. Furthermore, it was also not a special campaign day or holiday in the Netherlands. Despite the fact that certain minutes of the race in Monaco deviate, these minutes are still included because in these minutes the examined logos are often seen on average. This does not apply to Australia, where time was most likely to be the problem. On average, fewer people are on the website in the morning, what can be seen from the data. The time of the Miami race is not a time with a lot of website traffic, this is why the supermarket started a scraper at 10 p.m. to restart the data collecting. Data samples after 10 p.m. will not be taken into account for this reason. Yet 4487 rows remain, because no logos were seen on these times. For further research, it is important to keep in mind that there is a scraper that started at 10 p.m..

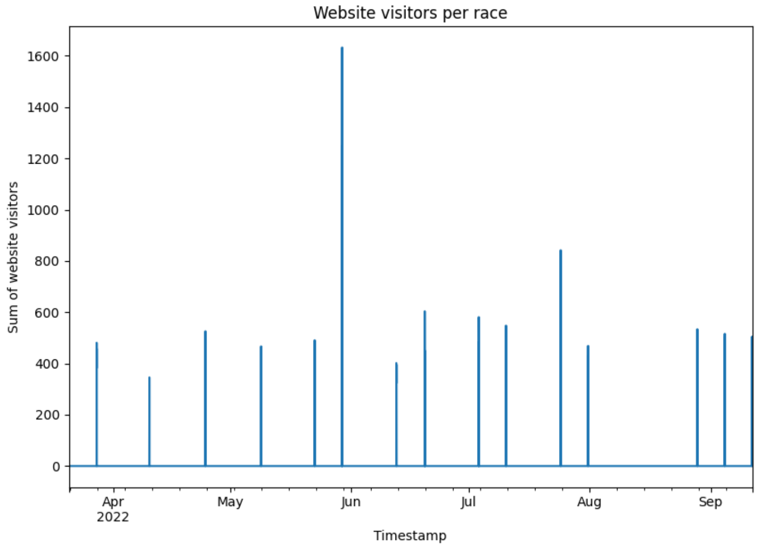

Figure 2: The number of website visitors per race for the entire period, after processing the data

Extra feature day parts

This feature is created to see if the races on specific parts of the day affect the number of website visitors. To measure this a correlation between these variables has been created. This is created to give a simple and quick summary of the strength of the relation. There is no need to predict variables, in this case a regression would be better. A strong correlation is close to 1 or -1 and a weak correlation is close to 0 (O'Brien & Scott, 2012). This extra feature, that is needed for the correlation is divided into three string categories: morning (06:00 a.m./noon), afternoon (12:01 p.m./6:00 p.m) and evening(6:01 p.m./midnight). As a result, a new dataset with 10 features and 750 samples is now created.

Data stationary

After generating and elaborating on the dataset, the difference between the features will be measured with different time series regression machine learning models. Before the analysis can be performed, some model preparation that applies for every model must be done. To start, loading in the data in the right way to get the timestamp as the index of the data. To see any trends or seasonality, functions and graphs were used to visualize the data before running the analysis. Some weekend days in the summer or winter and special days with campaigns can make a difference. Augmented Dickey-Fuller test is created to measure seasonality, this is displayed in table 1 The p-value is lower than 0.05 which means that the data is stationary. This can also be deduced from figure 2 where the complete timestamp of the data can be seen. There is no extreme increase or decrease in this period, except for a few outliers.

Test	T-Statistics	P-value	Critical value: 1 %	Critical value: 10%
ADT	-4.38	0.003144	-3.4317966	-2.5671102

Table 1: Results of Augmented Dickey-Fuller test, which indicate that the data is stationary

Supervised learning & normalization

When using time series algorithms, the time series data must be converted to supervised learning problem. Where the value(s) of x influence the y variable (y = f(x)). Extra lags of the y variable were created. Lagging a time series is moving timestamps forward or backwards. These were not created to make the data stationary, but to measure the data dependencies. To compare lags and see if there is any difference in the y variable (the website visitors). From the website visitor feature lags are created by taking the difference from the visitors and the timestamps, what results in the time categories, 1 minute before seeing the logos, 1 minute, 10 minutes and 1 hour after seeing the logos. The lags are in minutes because short-term behavior is measured. For all the algorithms, the lags are the same, to compare the results per algorithm.

To calculate the relationship in the data, scaling the data is necessary when multiple variables used. Because the scale can be different between the variable. When every variable is commonly scaled by their own distribution, the performance will often get a boost. The data is this study is stationary, no trends or seasonality. Normalization is chosen in this case, above Standardization. Besides that, a strong Gaussion distribution is recommended when using the Standardization technique (Asesh, 2022). The function Min-Max scaling belongs to normalizing the data, it re-scales it in values between 0 and 1. By using this formula (8), that subtracts the Xi of Xmin (input value minus the minimum of the input values) divided by Xmax minus Xmin (the maximum minus the minimal of the values).

$$Normalized(Xi) = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{8}$$

4.2.3 *Data split*

The dataset was split into three subsets: training, validation and test data. A random split of samples can lead to strange results, because an LSTM model uses the sequential of a dataset. This means that the training data consist of the first 70% of the data, the validation data is made up of the the next 15%, of the data and the last 15% are for the test data. They all have the same features, but with different timestamps. Training data is used for training the model. Validation data is created for tuning the hyperparameters and the test data is new data where the tuned model can be tested to see if the model can create results on new data. This balance is made to avoid underfitting and overfitting, and this split is commonly used in supervised learning. For the ARIMA and XGBoost the same balanced is used, since the models will be compared.

4.2.4 *Evaluation metrics*

To compare the algorithms on their prediction power, the following three evaluation techniques were performed. In time series analysis these three techniques are often used (Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018), (Reich et al., 2016), (Onyutha, 2020). Besides this, the metrics were also used to distinguish the best lags and the worst lags from the models. Mean absolute error (MAE) (9), root mean squared error (RMSE) (10) and R-squared (R2) (11) are used to perform this. Their formulas are the following, where x_i the observed x values are, y_i the observed y values are, n or N the sample size is and $\hat{x}_i$ the predicted values are:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i^{-} \hat{x}_i\right)^2}{N}} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(x_i^{obs-} \hat{x}_i^{pred}\right)}{\sum_{i=1}^{N} \left(x_i^{obs-} \hat{x}_i^{mean}\right)} \tag{11}$$

The MAE was used as evaluation criteria to see how large the error is between the predicted and the actual normalized value. MAE also works well with outliers. The RMSE is calculated on the same scale as the MAE, however, it punishes larger errors (outliers) more. This is important because a model that only predicts the mean may have a good MAE but not penalize outliers or bigger errors. In addition, it is also better if the value is closer to 0. Lastly, R2 is a good indication of how well the model explains variation in the data. An R2 closer to 1 indicates a model that describes the data well, whilst values that are closer to 0 or are negative will indicate a bad model fit.

4.3 *Hyperparameter tuning*

Before running ARIMA, LSTM and XGBoost and doing the analysis to get the results, hyperparameter tuning still needs to be done to increase the model fits. A few different techniques were used to generate the best parameters. This process is described below.

4.3.1 *ARIMA*

To pick the optimal parameter for the moving average (MA) p an autocorrelation function (ACF) plot is used, see figure 3. The graph starts with high values and decreases overtime, means that the correlation is low. The partial autocorrelation function (PACF) plot is created to find the Autoregressive values (AR), also called parameter q, see figure 4. It starts again with a spike and decreases overtime, but in this case there is a correlation between lag 1 and lag 2. Since the correlation is low between the lags in the ACF plot, p will be 1. And because of that small sign of auto regression started from lag 2, q is set to 1 as well. For the order of difference, parameter d. The value was set to 0, because the data was already stationary.

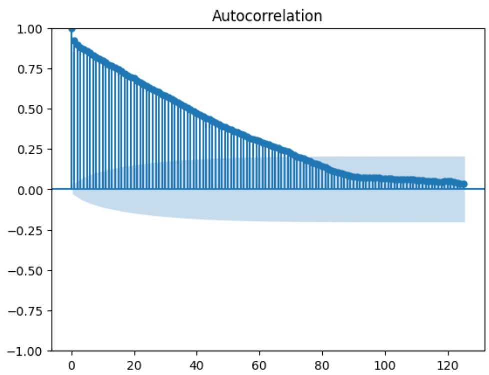

Figure 3: ACF plot,

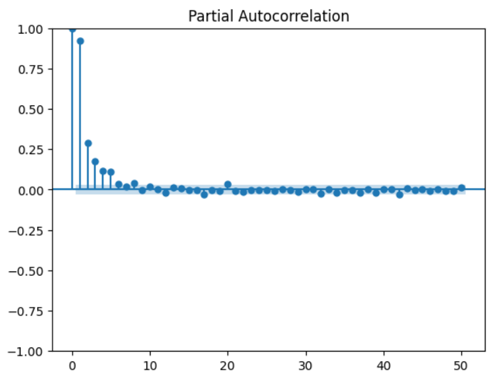

Figure 4: PACF plot

To control these results of the plots, autoarima is used for selecting the optimal hyperparameter p, q, d. Autoarima selects the best hyperparameters. There were several hyperparameters chosen to tune for ARIMA: seasonal (to take seasonal into account or not), max_p, max_d and max_q. These parameters have the largest impact in this model and were therefore changed from default. Seasonal is *False*, because there is no seasonal. The parameters p, q, and d were best when p=1, q=1 and d=0 (1,1,0).

4.3.2 *LSTM & XGBoost*

First the models will be trained on the training data, afterwards the hyperparameters settings will be tuned on the validation data, and the results will be tested on the test set. For LSTM and XGBoost grid search and experiments were used to find the best hyperparameters in order to get the best results. The following computer power properties were used to perform the gridsearch of LSTM: 96 vCPUs, 360 GB RAM NVIDIA T4 x 4. LSTM needs more computational power than e.g., XGboost. This is due to the use of a lot of memory and the sequential computation, to calculate the hidden layers. For LSTM, were the following hyperparameters have been tuned: the number of neurons, the batch size, the number of epochs, the learning rate, the optimizer, the loss function , the activation function and choosing the best training method, stateful or stateless. These search space of parameters is reduced by making specific choices based on the model. In addition to this parameter, there is also another function called Dense. Dense is set to 1 to create a single output layer of all the hidden layers. The number of neurons was set to 1. This number is based on experiments. The batch size was set to 41 and the epochs are optimal when they have a value of 195, with an early stopping patience value of 25 to prevent overfitting (Pasini, 2015). The learning rate (eta) shrinks the weights every step until it got the optimal variable (global minimum). The larger the eta, the more robust the steps are, which can cause missing the global minimum. The optimal eta is 0.15 for this data. This learning rate will be set for an optimizer like 'Nadam', 'Adam', 'RMSProp'. The Adam performed the best. Every step, there will be an estimate of the loss to update the weight to minimize the loss on the next step. For this LSTM, mean squared error is the best loss function. For the input and output layers, activation functions were used to decide if they were important or not during the prediction process. This Relu activation function fits best with this model. Finally, stateless resets the training process of the data every batch. Stateful goes on and the process will might be influenced by the previous step. For this LSTM, stateless has been chosen because the best batch size is higher than 1, and it uses the previous timestamp value which helps with predicting (Koudjonou & Rout, 2020).

The XGBoost hyperparameters were also tuned with grid search without more computational power. The following hyperparameters were tuned: max_depth, gamma, lambda, learning rate, eval metrics and n_estimators. In addition, because of the non-linear data, gbTree is used instead of gblinear. These remaining important parameters were changed from default when using a XGBoost time series. The max depth of the tree is 6, it is reducing overfitting. The gamma decides where the tree should split based on the minimum loss reduction of 0.1. Lambda handles the regularization part of XGBoost, 6 is the best weight of the L2 regularization. The optimal eta is 0.025 for this data. In the other models mean absolute error (MAE) is used to evaluate every step, because of the important outliers which have been taken. MAE can work better with outliers. For the XGboost, the loss function for this reason is MAE. The optimal number of trees (n_estimators) that were used in the model is 800. Every tree tries to minimize the error of the previous tree.

In figure 5, the process is displayed in a flowchart. From start to end, it is explained how this research is going to look.

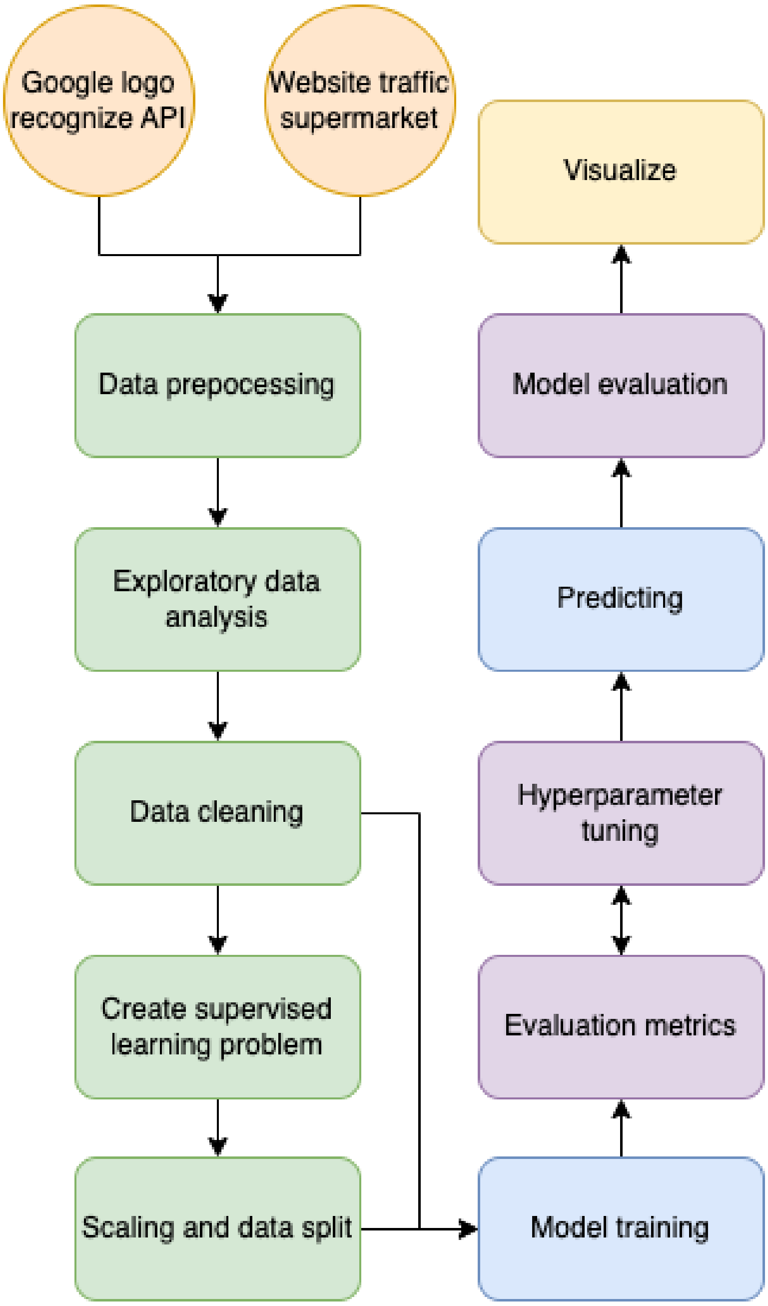

Figure 5: Process of the research, where every step of the analysis is described

4.4 *Packages*

This research was created by using Python version 3.10.7 (vanRossum, 1995) and Google BigQuery (Naidu & Tigani, 2014). The packages used for the analysis were the following:

- Pandas (McKinney, 2012)
- NumPy (Harris et al., 2020)
- Matpotlib (Hunter, 2007), seaborn (Waskom, 2021)
- XGboost (T. Chen & Guestrin, 2016)
- Autoarima (Melard & Pasteels, 2000)
- Keras (Géron, 2022)
- Tensorflow (Abadi et al., 2016)
- Scikit-learn (Pedregosa et al., 2011)

5 RESULTS

In this section, time series performance of the models on predicting website visitors are described and discussed. In the first section, the 'overall results' present the results of the tuned models and predictive performances. Afterwards, the next sections, 'ARIMA', 'LSTM' and 'XGBoost' present a more detailed analysis about the predictive power results per model, based on table 2 and the graphs. The best time category will be shown per model. The remaining results of the models are displayed in appendix A. Lastly, in the section 'importance day part' a correlation table is visualized that displays the correlation between the time categories and day parts. To illustrate the importance of a day part on the website visitors fluctuation. An overall note, the data in the models are scaled and the x axes indicate the and entire timestamp, which starts from the 19th of March and last until the 11th of September. In addition, MAE is the most important loss metrics in this study, because it can deal best with the outliers (most robust). This means that the MAE could be decisive as a result.

5.1 *Overall results*

First, the model losses RMSE, R2 and MAE of the ARIMA, LSTM and XGBoost on the y variable 'website visitors' are compared. Compared with the time series categories of 1 minute before (baseline), 1 minute, 10 minute and 1 hour after seeing the logo, this table with the overall results of the models can be seen in 2. The values in bold are the best results.

Table 2: The overall results of the analysis, where the models and time categories are compared based on the evaluation metrics

Timestamp	Evaluation metrics	**ARIMA**	**LSTM**	**XGBoost**
1min before	RMSE	0.562	0.449	0.545
1min after	RMSE	0.478	0.394	0.387
10min after	RMSE	0.406	0.365	0.441
1 hour after	RMSE	0.384	**0.354**	0.630
1min before	R2	-0.609	-3.376	-0.350
1min after	R2	-0.542	-2.728	-0.014
10min after	R2	-0.006	-1.583	-0.19
1 hour after	R2	**0.148**	-1.786	-1.305
1min before	MAE	0.448	0.398	0.440
1min after	MAE	0.377	0.380	0.312
10min after	MAE	0.319	0.335	0.354
1 hour after	MAE	**0.276**	0.301	0.513

5.2 *ARIMA*

As can be seen in figure 6, the number of website visitors after one hour can be predicted best. The model can predict the time series very well, it fits the line of the actual scaled values close to perfect. It out-performed the baseline. In addition, scores the ARIMA model and this time categories on almost every evaluation metrics the best, what can been seen in table 2. Based on the best timestamp, it predicts as second best; data with outliers, with a scaled RMSE score of 0.384 (closest to 1). This model can explain the variance of the data best, with a scaled R2 of 0.148 (closest to 1) and it also performed best when there are no outliers, it got a scaled MAE of 0.276 (closest to 0). This one hour time category is followed by 10 minutes, 1 minute after the logos, and the baseline 1 minute before seeing the logos, the baseline performed worst with a RMSE of 0.562, a R2 of -0.609 and a MAE of 0.448. These model fit graphs can be found in appendix A *-Results ARIMA*.

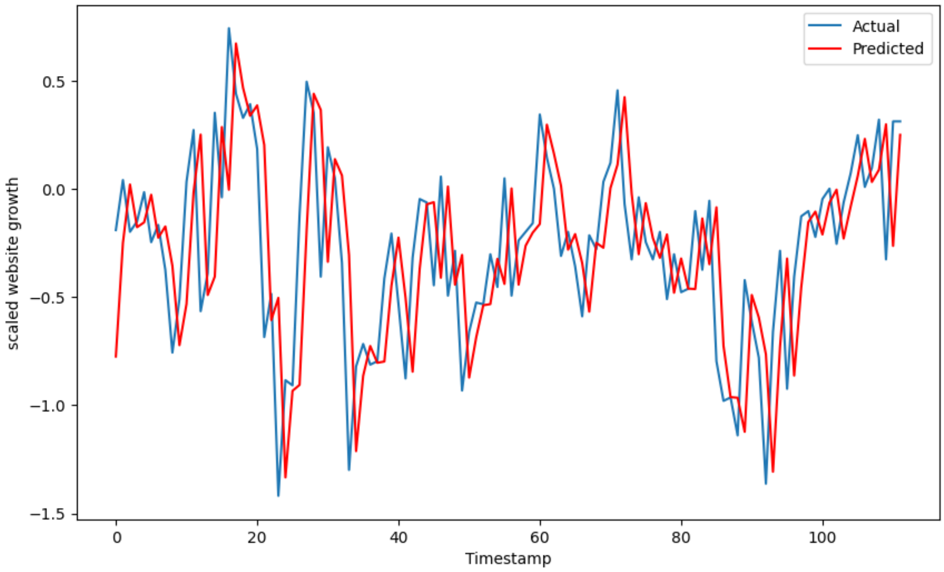

Figure 6: The predicted scaled website visitors versus the actual scaled website visitors, one hour after seeing the logo (ARIMA)

5.3 *LSTM*

LSTM can predict the time series second best behind the ARIMA and ahead of XGBoost. The prediction line fits the actual scaled values not that well, displayed in figure 7. But is better than the baseline. Predicting one hour after seeing the logo performs generally best here as well, with a RMSE of 0.354, a R2 of -1.786 and a MAE of 0.301 what is displayed in table 2. Followed by 10 minutes, 1 minute before seeing the logos and 1 minute after seeing the logos, these results graphs can be found in appendix A *-Results LSTM*.

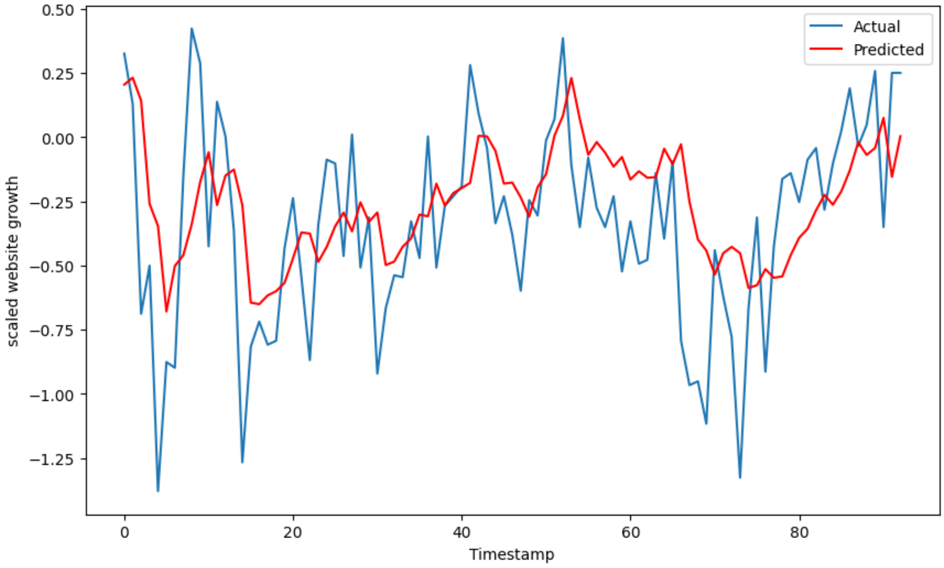

Figure 7: The predicted scaled website visitors versus the actual scaled website visitors, one hour after seeing the logo (LSTM)

5.4 *XGBoost*

Based on table 2 and the graphs from the other models, there can be concluded that XGBoost predict this data as worst, see figure 8 The predictions outcomes of the time category differ from the other models. The best time category is for XGBoost the 1 minute after seeing the logo. This category got a RMSE of 0.387, a R2 of -0.014 and a MAE of 0.312. The predictions of 10 minute after seeing the logo scores second best, followed by the baseline (1 minute before seeing the logos). The worst performance is the category 1 hour after seeing the logo, that got a RMSE of 0.630, a R2 of -1.305 and a MAE of 0.513. These less performing results can be seen in Appendix A -*Results XGBoost*.

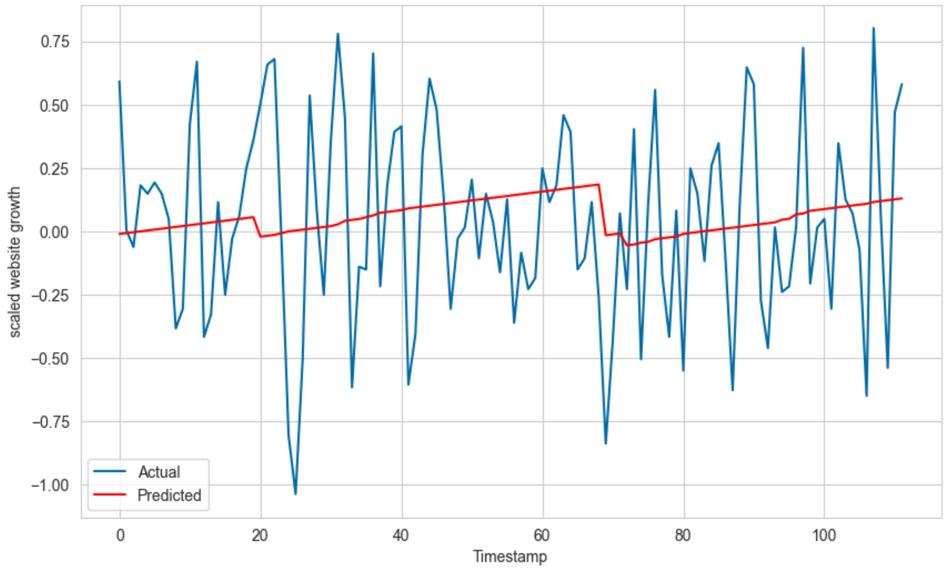

Figure 8: The predicted scaled website visitors versus the scaled actual website visitors, 1 minute after seeing the logo (XGBoost)

Based on the figures 7 and 8, there can also be seen that the most loss spikes can be found by the small outliers. The model does not understand what to do with the small outliers, because they do not occur often due to the low number of special days or events on these Sundays. Therefore it make sense that these are the worst predicted. These outliers peaks are throughout the year, so it is difficult to say whether there is still seasonality in the data.

5.5 *Logo comparison*

In this paragraph the influence of the chosen logos will be measured. The logo of the supermarket and Red Bull are compared to create discussion points regarding the influence of a specific logo during a Formula One race. For this subset comparison the ARIMA model and the time categories of '1 hour after seeing the logo' is used, because of their performance. This time category is compared to the previously used baseline. In contrast, the same evaluation metrics are used for these comparisons. Based on table 3, the results for the Red Bull logo were better looking at the baseline category and the used time category of the logos. The used time category had a RMSE of 0.415, R2 of -0.123 and a MAE of 0.323. It does make sense, the results do not differ much from table 2, because the larger part of the dataset consist out of Red Bull logos.

Table 3: The comparison in performance between the logos, which where used in the analysis

Company		Supermarket		Red Bull	
		1 min before (baseline)	1 hour after	1 min before (baseline)	1 hour after
ARIMA	RMSE	0.801	0.704	0.634	**0.415**
	R2	-1.101	-1.696	-0.674	**-0.123**
	MAE	0.697	0.599	0.469	**0.323**

5.6 *Importance day parts*

In this section the influence of day parts on website visitors is measured and featured in table 4. The numbers in the table display the power of the correlation between the website visitors per time category and day parts. In other words does additional number of visitors occurs when the formula one race is held on a specific part of the day, like the morning, afternoon or night. It shows that none of the day parts is in correlation with the time categories, because they are close to 0. The most correlated time category is 1 hour after seeing the logos with a correlation of 0.18. Which still means that a day part does not affect the number of visitors 1 hour after seeing the logos. The minute after seeing the logo performed as the worst, with a correlation of 0.014. The baseline is not taken into account, it does not tell anything about the number of visitors based on the influence of the logos. Based on these results, no further research of the categories, morning, afternoon or night has been done.

Table 4

Correlation			
	1 min after	10 min after	1 hour after
Day parts	0.014	0.056	**0.18**

6 DISCUSSION

The research goal for this thesis is to measure how accurately the website visitors can be predicted by using the commercials during a Formula One race. The sub-questions suggest a comparison between models, that are the machine learning models: ARIMA, LSTM and XGBoost. In addition, a comparison is made between different time categories, which influence the website visitors the most. These categories consist the website visitors 1 minute, 10 minutes and one hour after seeing the logos and 1 minute before seeing the logos, this is the baseline. Finally, the day parts when races where held was also investigated. Where the morning, afternoon and evening were compared on the different time categories. The data sets retrieved from Springbok (2016) and Google API were used to make these comparisons. The results displayed that the ARIMA model with the time category, 1 hour after seeing the logo was the best. With a RMSE of 0.562, a R2 of 0.148 and a MAE of 0.276. This time category also came out best in the correlation when examining dayparts, with a correlation of 0.18.

6.1 *Results discussion*

Results showed that the ARIMA model in combination with the time category 'one hour after seeing the logo' predicted the website visitors best. The ARIMA generated for this time category the lowest error (RMSE, MAE), that is the closest to 0. And the best model fit (R2), which is the closest to 1. However, these are still not the optimal results for the outcome of a research.

A lot of researchers Li and Zhang (2018), Siami-Namini and Namin (2018) showed that the LSTM model often out-preforms the ARIMA model when predicting time series. For this part, reasons can be created that the data set is processed to a univariate dataset before every new time category run. This might made the data to simple for the other models, but this is actually good for the ARIMA (Shelatkar et al., 2020). What also was not helpful for the other models is the size of the dataset. In addition, the XGBoost performs considerably worse, Carreón et al. (2019) concluded the same for the XGBoost, when predicting the influence from commercials on short-term behavior.

On the other hand, two of the three models showed that there is a gradient from short-term to longer-term behavior is. Sama (2019) researched this topic and concluded the same as this study, that it is important to give customers a lot of small reminders to change their short-term behavior. This also ties in with the logo comparison. Red Bull logos have appeared more often during sporting events, which has led to better results compared to the supermarket logo.

Avraam et al. (2021) showed that there is a peak of online visitors between 6:00 p.m. and 10:00 p.m.. This study shows the opposite, there is no correlation with a specific day part. So the number of website visitors are not affected by this peak.

Interestingly, the most outliers were also mentioned in the analysis, what created noise in the normal distribution of the data. Some of these outliers were hard to predict for the LSTM and XGBoost models. In contrast, the ARIMA model handled these outliers well. Which may also be due to the univariate and simplicity of the data.

6.2 *Limitations*

It cannot be said concretely that website traffic is influenced by sponsoring at sports events. Despite the delimited scope, there are still many other confounding influences that can drive people through a website visit, for example, offline marketing.

Given the dataset, we could only study the relationship of 750 rows from 16 races where a logo appeared. In the rest of the race minutes there did not appear a logo of the supermarket or Red Bull. With more data, more minutes with logos could have been explored that might prove more influence on the website visitors. Also, the size of the dataset and the type of data has limited the modeling options to be used. Specific time series machine learning technique have proven to be successful in predicting time series and website traffic, however, they require a larger dataset or more features. A univeriate dataset is used as input for the models, for some models is it more difficult to work with this simple dataset form.

6.3 *Future research*

This thesis represents a step toward a field of study with numerous opportunities. It approached a topic with predictive modeling that has not often been tested in this way before. It prioritizes the predictive implications rather than the explanation of the commercials effects that influence the website visitors. Using more often and other predictive implications might increase the effect of commercials on website visitors.

Also, this research can be extended to more companies that are running commercials during a Formula One race. Now, are the detected logos related to this specific supermarket. When exploring different companies, especially those that have a large history of running commercials during sports events, might discover new patterns. In addition, perhaps other types of sport or adding more logos increase the results, and might show the difference in type of sports or logos. Which might prompt companies to change there sponsorships or logo.

7 CONCLUSION

The aim of this study is to detect the number of website visitors created by television commercial at sport events and if the day parts when the race is held influence the number of visitors. This research is created based on the questions arising around the return of investments of sponsorships in the supermarket branch. Due to the increase in online traffic and the ignorance about this return. The following questions are created based on this gap and to fulfill these questions:

RQ1 *which of the researched time series machine learning models can predict the number of website visitors most accurately?*

This sub-question aims to predict in the most accurate way, the website visitors that are influenced by the commercials. For this predictions the following time series models were used; ARIMA, LSTM and XGBoost. There can be concluded that the best model for predicting the website visitors is ARIMA model. Which got the lowest error rates in general.

RQ2 *Which of the researched time categories can be used to predict the website behavior best?*

This sub-question aims to select the best time category that got the lowest error rate when predicting the number of website visitors. The following categories of short-term behavior are predicted, 1 minute before seeing the logo, 1 minute and 10 minute after seeing the logo. And 1 hour after seeing the logo falls under long-term behaviour in this study. Based on the results there can be concluded, that the long-term behavior, '1 hour after seeing the logo' is the best time category to use for the prediction.

RQ3 *To what extent do day parts influence the number of website visitors prediction when the sports event is happening?*

The last sub-questions aims to find a correlation between the parts of a day a race is held and the growth of website visitors. The highest correlation was 0.18, what results in a weak correlation. There can be concluded that the specific day part does not influence the growth of the website visitors.

MQ *To what extent can the number of website visitors be predicted, based on television commercials on a specific day part during sports events?*

It can be concluded that the supermarket or other companies should use the ARIMA model in the following predictions of website usage by commercials at sports events. This model provides the best results overall. The time category, '1 hour after seeing the logos' generated the lowest errors. Long-term behavior will be more influenced than short-term behavior.

However, the results are not reliable because it got scaled errors of 0.562 (RMSE), 0.148 (R2) and 0.276 (MAE). The researched category, 'day parts' does not influence the number of website visitors. So, the specific time a race is held is not important. Since the quantity and type of data was a limitation of the thesis. Subsequently, research must be done on different data sets and for different types of sports which additional functions can be added to the analysis.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In *12th usenix symposium on operating systems design and implementation (osdi 16)* (pp. 265–283).

Abbasi, R. A., Javaid, N., Ghuman, M. N. J., Khan, Z. A., Ur Rehman, S., et al. (2019). Short term load forecasting using xgboost. In *Workshops of the international conference on advanced information networking and applications* (pp. 1120–1131).

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, *8*(1), 1–74.

An, H., & Ren, J. (2020, 04). Xgbdeepfm for ctr predictions in mobile advertising benefits from ad context. *Mathematical Problems in Engineering*, *2020*, 1-7. doi: 10.1155/2020/1747315

Asesh, A. (2022). Normalization and bias in time series data. In *Conference on multimedia, interaction, design and innovation* (pp. 88–97).

Avraam, E., Veglis, A., & Dimoulas, C. (2021). Publishing patterns in greek media websites. *Social Sciences*, *10*(2), 59.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, *29*, 105340.

Cameron, N. (n.d.). Understanding sponsorship and its measurement implications. *Journal of Sponsorship*, *2*(2), 131–139.

Carreón, E. C. A., Nonaka, H., Hentona, A., & Yamashiro, H. (2019). Measuring the influence of mere exposure effect of tv commercial adverts on purchase behavior based on machine learning prediction models. *Information Processing & Management*, *56*(4), 1339–1355.

Casado-Vara, R., Martin del Rey, A., Pérez-Palau, D., de-la Fuente-Valentín, L., & Corchado, J. M. (2021). Web traffic time series forecasting using lstm neural networks with distributed asynchronous training. *Mathematics*, *9*(4), 421.

Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C.-H., & Liu, R. (2019). Xgboost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access*, *7*, 13149–13158.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Cobbs, J. B., Jensen, J. A., & Tyler, B. D. (n.d.). The sponsorship performance cycle: longitudinal evidence of sponsors' contribution to formula one team achievement. *Journal of Business & Industrial Marketing*, *37*(1), 103–114.

Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, *74*, 902–924.

Dolega, L., Rowe, F., & Branagan, E. (2021). Going digital? the impact of social media marketing on retail website traffic, orders and sales. *Journal of Retailing and Consumer Services*, *60*, 102501.

Fanoodi, B., Malmir, B., & Jahantigh, F. F. (2019). Reducing demand uncertainty in the platelet supply chain through artificial neural networks and arima models. *Computers in biology and medicine*, *113*, 103415.

Farzad, A., Mashayekhi, H., & Hassanpour, H. (2019). A comparative performance analysis of different activation functions in lstm networks for classification. *Neural Computing and Applications*, *31*(7), 2507–2521.

Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. " O'Reilly Media, Inc.".

Grohs, R. (2016). Drivers of brand image improvement in sports-event sponsorship. *International Journal of Advertising*, *35*(3), 391–420.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., . . . others (2020). Array programming with numpy. *Nature*, *585*(7825), 357–362.

Hawkins, D. M. (1980). *Identification of outliers* (Vol. 11). Springer.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(03), 90–95.

Jensen, J. A., & Cobbs, J. B. (2014). Predicting return on investment in sport sponsorship: Modeling brand exposure, price, and roi in formula one automotive competition. *Journal of Advertising Research*, *54*(4), 435–447.

Kakalejčík, L., Bucko, J., & Danko, J. (2020). Impact of direct traffic effect on online sales. *Journal of Research in Interactive Marketing*, *14*(1), 17–32.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied soft computing*, *11*(2), 2664–2675.

Koronios, K., Ntasis, L., Dimitropoulos, P., & Ratten, V. (2021). Not just intentions: predicting actual purchase behavior in sport sponsorship context. *Sport, Business and Management: An International Journal*.

Koudjonou, K. M., & Rout, M. (2020). A stateless deep learning framework to predict net asset value. *Neural Computing and Applications*, *32*(14), 1–19.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018).

Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022.

Li, P., & Zhang, J.-S. (2018). A new hybrid method for china's energy supply security forecasting based on arima and xgboost. *Energies*, *11*(7), 1687.

Luo, J., Zhang, Z., Fu, Y., & Rao, F. (2021). Time series prediction of covid-19 transmission in america using lstm and xgboost algorithms. *Results in Physics*, *27*, 104462.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.

McKinney, W. (2012). *Python for data analysis: Data wrangling with pandas, numpy, and ipython*. " O'Reilly Media, Inc.".

Melard, G., & Pasteels, J.-M. (2000). Automatic arima modeling including interventions, using time series expert software. *International Journal of Forecasting*, *16*(4), 497–508.

Naidu, S., & Tigani, J. (2014). *Google bigquery analytics*. John Wiley & Sons.

Nielsen, D. (2016). *Tree boosting with xgboost-why does xgboost win" every" machine learning competition?* (Unpublished master's thesis). NTNU.

Ogunleye, A., & Wang, Q.-G. (2019). Xgboost model for chronic kidney disease diagnosis. *IEEE/ACM transactions on computational biology and bioinformatics*, *17*(6), 2131–2140.

Onyutha, C. (2020). From r-squared to coefficient of model accuracy for assessing" goodness-of-fits". *Geoscientific Model Development Discussions*, 1–25.

O'Brien, D., & Scott, P. S. (2012). Correlation and regression. *Approaches to quantitative research: A guide for dissertation students*, 106–124.

Pasini, A. (2015). Artificial neural networks for small dataset analysis. *Journal of thoracic disease*, *7*(5), 953.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Reich, N. G., Lessler, J., Sakrejda, K., Lauer, S. A., Iamsirithaworn, S., & Cummings, D. A. (2016). Case study in evaluating time series prediction models using the relative mean absolute error. *The American Statistician*, *70*(3), 285–292.

Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing*, *323*, 203–213.

Sak, H., Senior, A. W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. , 5.

Sama, R. (2019). Impact of media advertisements on consumer behaviour. *Journal of Creative Communications*, *14*(1), 54–68.

Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (arima) models: a guide for evaluating large-scale health interventions. *BMC medical research methodology*, *21*(1), 1–12.

Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A., & Vahdat, S. F. (2011). Modeling of weather parameters using stochastic methods (arima model)(case study: Abadeh region, iran). In *International conference on environment and industrial innovation, ipcbee* (Vol. 12).

Shelatkar, T., Tondale, S., Yadav, S., & Ahir, S. (2020). Web traffic time series forecasting using arima and lstm rnn. In *Itm web of conferences* (Vol. 32, p. 03017).

Siami-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: Arima vs. lstm. *arXiv preprint arXiv:1803.06386*, 19.

Smagulova, K., & James, A. P. (2019). A survey on lstm memristive neural network architectures and applications. *The European Physical Journal Special Topics*, *228*(10), 2313–2324.

Springbok. (2016). Retrieved from https://www.springbokagency.com/en/

vanRossum, G. (1995). Python reference manual. *Department of Computer Science [CS]*(R 9525).

Voorveld, H. A., Araujo, T., Bernritter, S. F., Rietberg, E., & Vliegenthart, R. (2018). How advertising in offline media drives reach of and engagement with brands on facebook. *International Journal of Advertising*, *37*(5), 785–805.

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Yang, Y., & Zhai, P. (2022). Click-through rate prediction in online advertising: A literature review. *Information Processing & Management*, *59*(2), 102853.

Yu, R., Gao, J., Yu, M., Lu, W., Xu, T., Zhao, M., ... Zhang, Z. (2019). Lstm-efg for wind power forecasting based on sequential correlation features. *Future Generation Computer Systems*, *93*, 33–42.

Zhou, K., Wang, W., Huang, L., & Liu, B. (2021). Comparative study on the time series forecasting of web traffic based on statistical model and generative adversarial model. *Knowledge-Based Systems*, *213*, 106467.

8 APPENDIX A - RESULTS MODELS

Results ARIMA

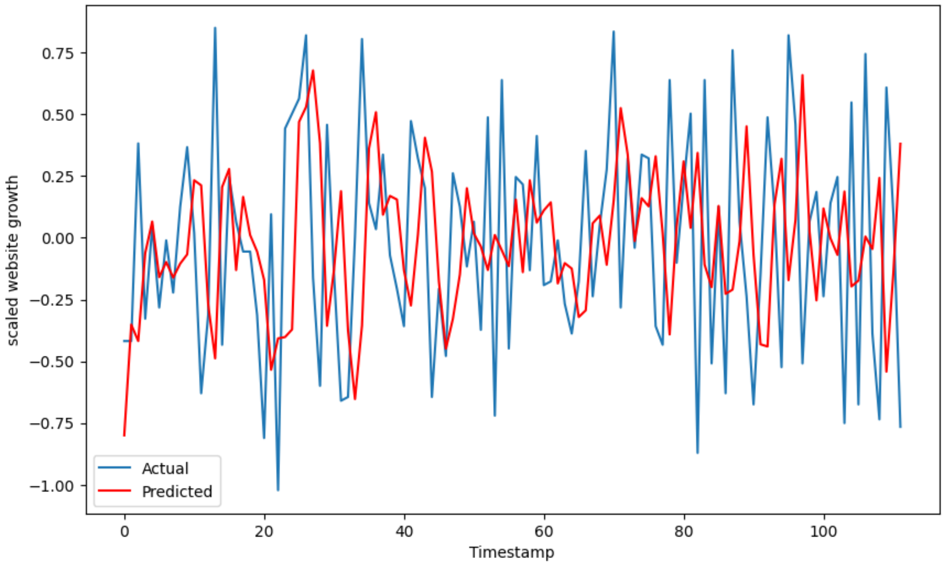

Figure 9: The predicted scaled website visitors versus the actual scaled website visitors, 1 minute before seeing the logo (ARIMA)

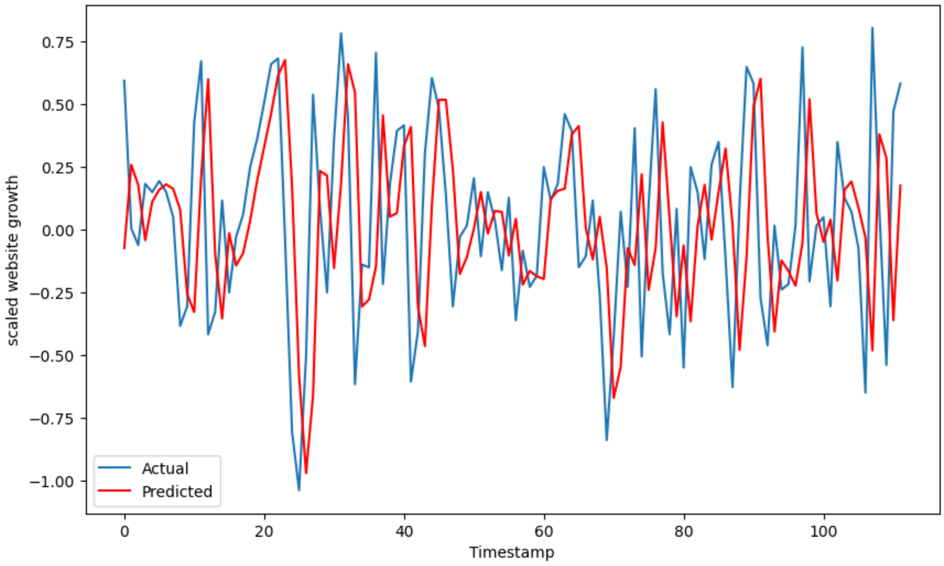

Figure 10: The predicted scaled website visitors versus the actual scaled website visitors, 1 minute after seeing the logo (ARIMA)

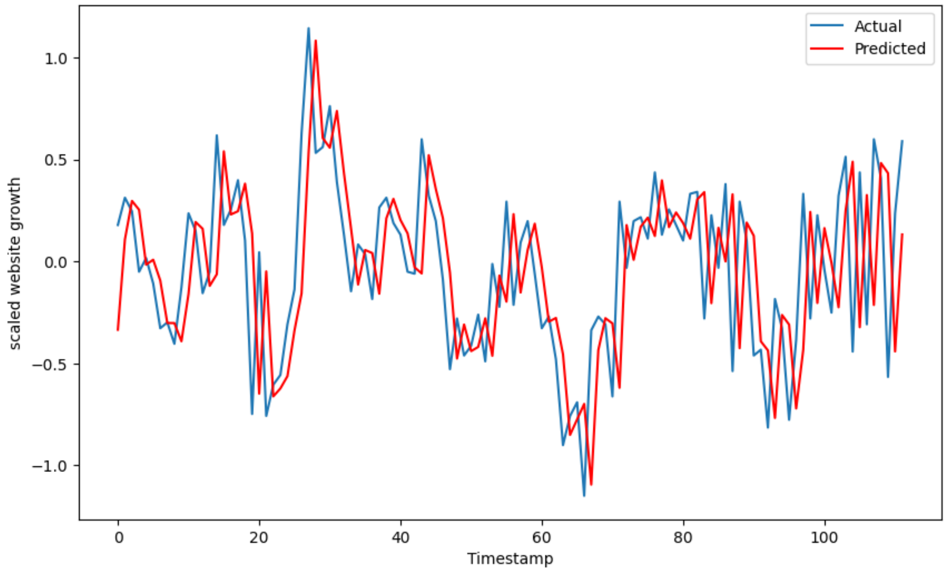

Figure 11: The predicted scaled website visitors versus the actual scaled website visitors, 10 minutes after seeing the logo (ARIMA)

Results LSTM

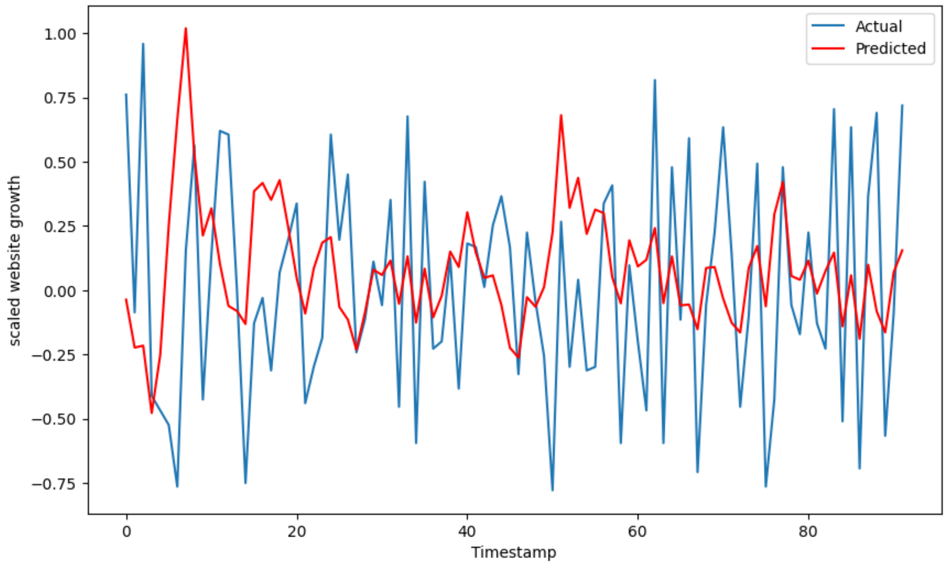

Figure 12: The predicted scaled website visitors versus the actual scaled website visitors, 1 minute before seeing the logo (LSTM)

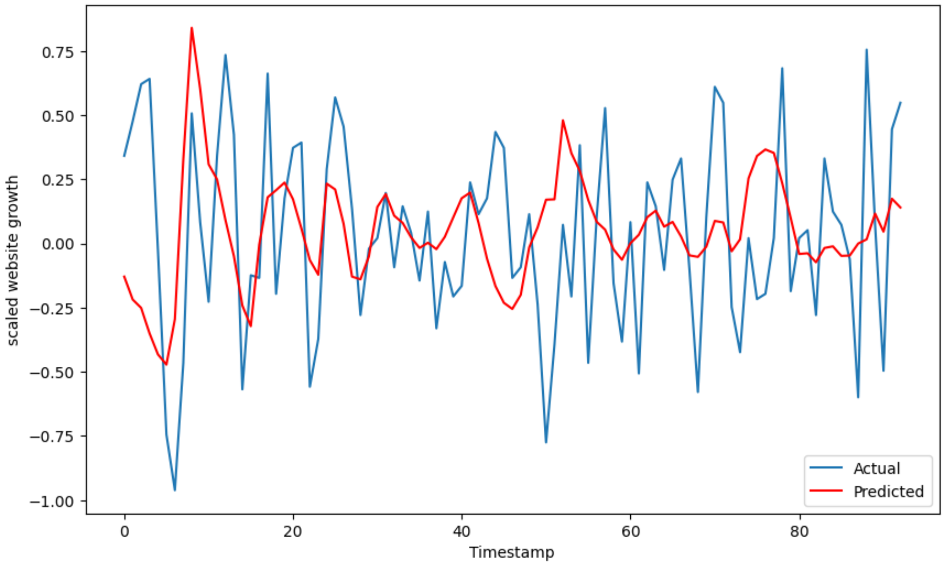

Figure 13: The predicted scaled website visitors versus the actual scaled website visitors, 1 minute after seeing the logo (LSTM)

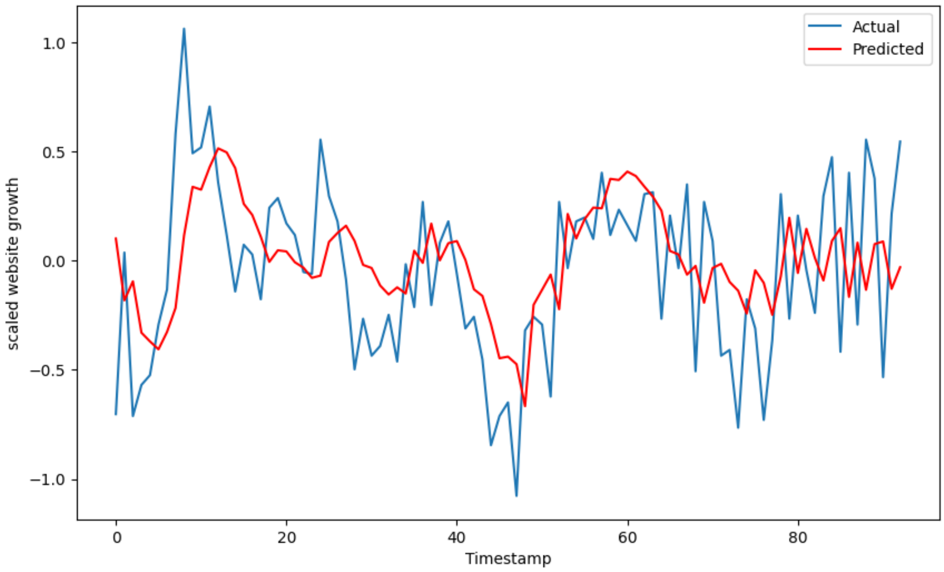

Figure 14: The predicted scaled website visitors versus the actual scaled website visitors, 10 minutes after seeing the logo (LSTM)

Results XGBoost

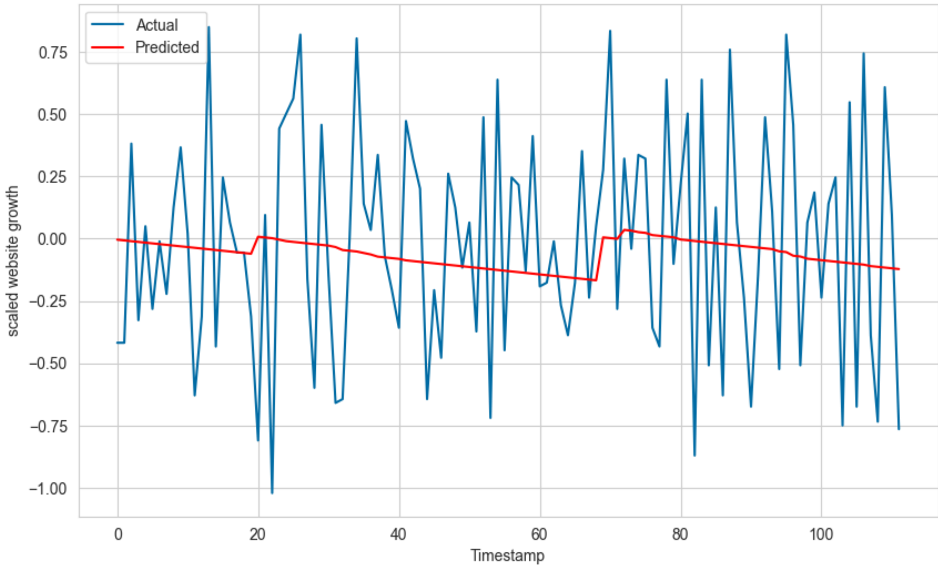

Figure 15: The predicted scaled website visitors versus the actual scaled website visitors, 1 minute before seeing the logo (XGBoost)

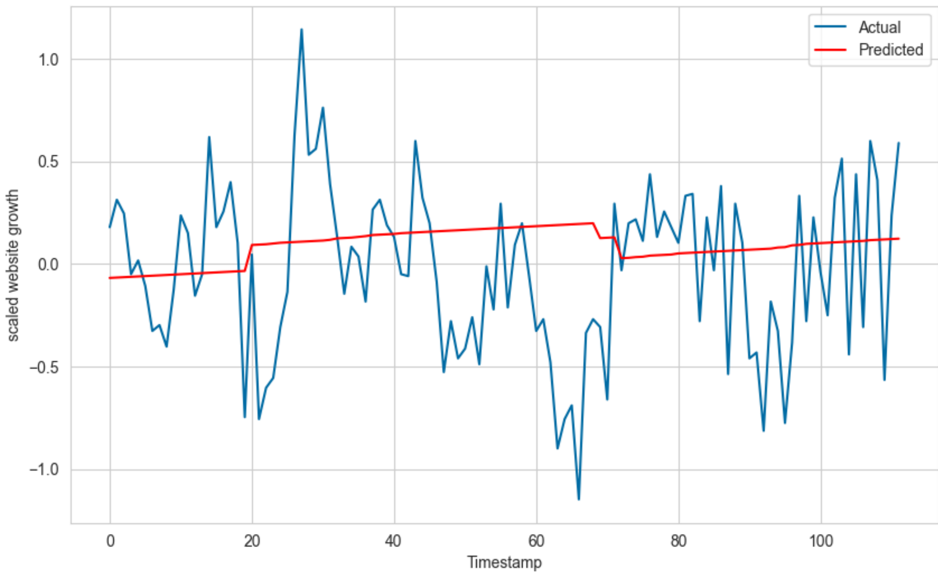

Figure 16: The predicted scaled website visitors versus the actual scaled website visitors, 10 minutes after seeing the logo (XGBoost)

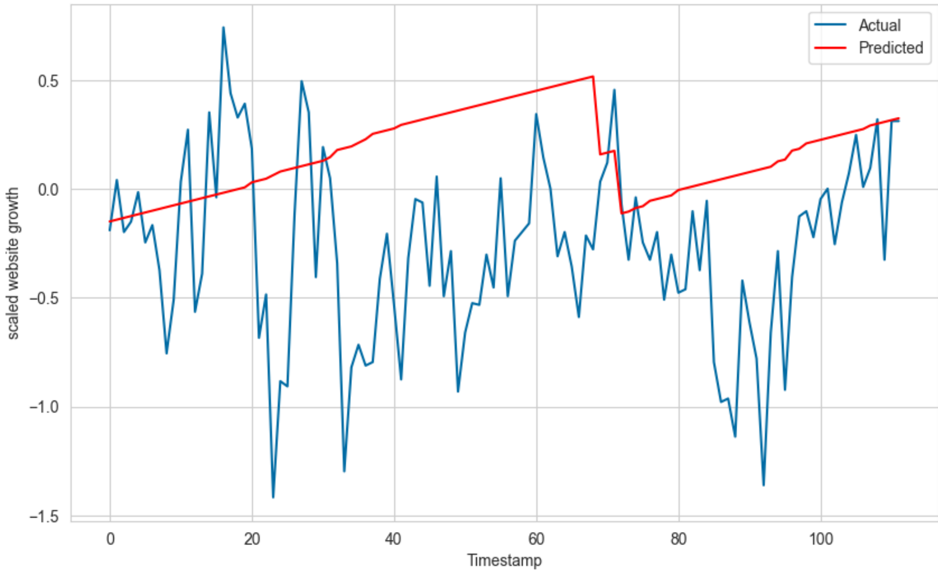

Figure 17: The predicted scaled website visitors versus the actual scaled website visitors, 1 hour after seeing the logo (XGBoost)