Universiteit Leiden

ICT in Business and the Public Sector

Explaning Market Share Prediction

A Study On the Use of Machine Learning in E-Commerce to Predict Market Share and the Connected Implementation of Explainable Artificial Intelligence Techniques to Increase Interpretability of these Predictions for Business Stakeholders with a Non-Technical Background

Name: Nicole van Halm
Student-no: s2992574

Date: 02/08/2022

1st supervisor: Dr. M. Hilbert
2nd supervisor: Prof.dr.ir. J.M.W. Visser

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract. Companies are interested in knowing their market share, as it shows both their market power and has a positive relationship with profit. However, they often do not know what actions they can undertake to increase this market share. This study researched the use of Explainable Artificial Intelligence (XAI) to uncover features that are most important in market share prediction. To do so, the study experimented with a variety of supervised regression models to predict market share. The Light Gradient Boosting Machine model was chosen, based on its performance. Thereafter, multiple XAI methods were implemented to explain how the model made its predictions, showing the featurest that are most important. Business stakeholders responded with enthusiasm to the resulting explanations, stating that these would be useful within their work. The study found that within their responses to XAI, the stakeholders showed differing preferences in type of explanation. This could be explained by considering the similarity between the preferred XAI method and a participant's way of working.

Keywords: Machine Learning · Tree-Based Modelling · E-Commerce · Explainable Artificial Intelligence · Market Share · Stakeholder Understanding.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Problem Statement

Traditionally, market share is a metric used for brands, aiming to measure the effectiveness of their marketing efforts in a set location during a set time period [24]. However, recently the interest in knowledge on market share is also found in several Dutch e-commerce companies. This is reflected in the clientele that market research company such as Nielsen and GFK have acquired, including companies such as Wehkamp, TicketMaster, HEMA and Gorillas [36,68]. Nevertheless, there are not many models that have been developed to fit this new way of considering market share. This makes it difficult for a company to predict their upcoming market share, while this prediction could provide them with insights on important factors and whether they will perform well under certain conditions. The modelling of market share can be achieved with the use of Machine Learning (ML). Nevertheless, ML models could still fail to provide an understanding of market share in the way that is required to enable usage in the business.

This failure to provide understanding is related to the way that models are built. Currently, a lot of machine learning (ML) models in the field of Artificial Intelligence (AI) suffer from the black box problem. This problem refers to the lack of understanding that humans have of the processes within these models, leading to little comprehension of how model outputs have been reached [29]. While it is clear what the input and the output of a model is, why a model makes certain decisions or how it processes information can often not be found with just the model alone [100].

To counter the black box problem, many techniques have been developed in the field of Explainable Artifical Intelligence (XAI). With these techniques, XAI aims to explain the rationale behind the decision-making process of ML models, in order to define the (dis)advantages of a process and help predict the model's future outputs [76]. An important concept that these techniques focus on is *understandability*. Understandability (also known as intelligibility) refers to a characteristic of a ML model, which helps humans understand how the model functions without needing to disclose the algorithmic structure of said model [5]. Hence, the degree of understandability of a model can depend on the human audience that has the need to understand it. This makes the perception and reception of the audience trying to understand the model a determining factor [56]. In the cases where the human audience cannot grasp the ML models due to their complexity or their inherent opaqueness, insights from XAI become especially relevant [52].

There are different ways of implementing XAI and increasing a model's interpretability, depending on the anatomy of the model that needs understanding. The division of XAI methods is one that's made over phases: pre-model, in-model and post-model. One of the criteria for correctly choosing a method for a model is to look at whether the model has any intrinsic

interpretability. If so, this means that the model has attributes that makes it interpretable, showing how the model takes decisions [17].

However, neither the increased focus on XAI in research nor the development of many tools useful for XAI implementation, have yet been translated into many organizations' actual applications [56]. Among one of the areas in which this translation is not yet prominent is that of e-commerce. Nevertheless, in other fields that are related to e-commerce some XAI implementations can be found, especially in regard to customer experiences. Examples of this are (i) research done into XAI for customer churning in the context of banking and (ii) XAI to determine customer preferences in case of online advertising [15,77]. These studies concluded that their results seem promising but, due to the restricted scope of the conducted study, the generalization possibilities of the result are limited [77]. However, these studies do suggest that the application of XAI in the e-commerce context is possible, and academically relevant.

In consideration of the existing literature and the posed problem statement the following research question was decided upon: **How can XAI contribute to the prediction of market share in the field of e-commerce?**

1.2 Research Context

The aim of this section is to shed light on the business context in which the current study has been conducted.

Company
The formed research question will be answered within the context of the business environment of the company that has hosted this thesis. Being founded in 2003, the host company is the first online bookstore of the Netherlands. Even after expanding its product range to a broad selection, books remained an important part of the business. However, in 2020 another big e-commerce player has entered the Dutch market, leading to a decrease in the company's market share. Where the company held almost all of the sales done in the online book market before this entry, they were suddenly confronted with a significant shrinkage of this percentage.

Goal
As a response to this threatening newcomer the company's pricing team has invested into automatically updating its prices on a daily basis. Furthermore, measures in other departments, such as logistics and supply chain, have been taken to regain market share. However, it is unknown to the company to what extent each of these have had the desired effect. They wish to gain insight on what potential measures could help them increase their market share in the future. The goal of this research was to find important influencing factors of market share, in order to point the company in the direction of measures connected to these factors, helping the company reach their set market share target.

Scope
This research will focus on the online market share of international and educational international books. International books are books published in any language other than Dutch. Educational books are to be understood as books published with the intention to be used by students [1]. Dutch books are excluded as they are more difficult to influence because their prices are fixed by law [69]. For this reason the company does not consider them to be a focus group in terms of market share. References to the Dutch Book Market from now on will mean only those books that are covered by the defined scope.

Market Dynamics
As can be deduced from the explanation on the company's position, there are not many players in the Dutch online book market. This limited amount leads to those players holding a lot of influence over the market. In my experience at the host company I have noticed the following. If one competitor changes a price or promotes a certain book, the other(s) often follow in their footsteps. Furthermore, if a book performs well online this often means physical book stores will 'pick up' this specific book and start selling and/or promoting it themselves. In general, this indicates that online booksellers hold a lot of power over the book market in general.

Furthermore, a big difference between online and physical book stores is the amount of different books they can sell. Where physical stores only have limited space in their shops, online parties do not have this limitation. Often, online booksellers are not even bound to their own warehouse storage, due to the popularity of cross-docking and printing on demand. The less popular a book is, the more frequent these processes are used. Cross-docking generally occurs for 40% of the company's sales. The almost infinite amount of books that can be sold means that the Dutch online book market has a very big long tail, as shown in figure 1 [2].

[1] One is considered to be a student if one fulfills either the requirement of following any higher education or if one is following a (self-study) training/course for the purpose of learning from it.

[2] Please note that the maximum amount of items per book sold in the market reaches far higher than the maximum shown in the graph, this is a zoomed in image with the purpose to illustrate the long tail.

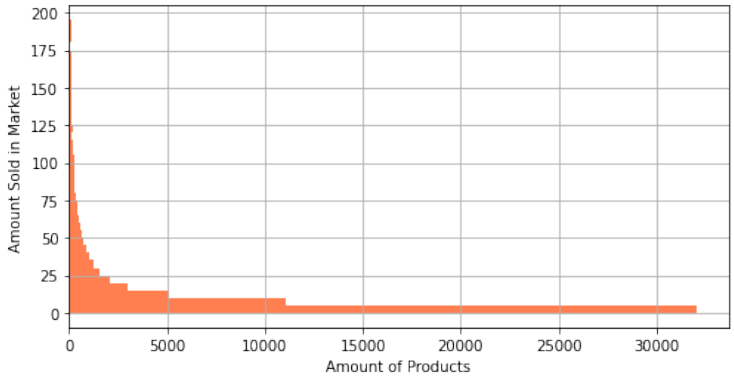


Fig. 1: Book Market Sales Distribution

1.3 Outline

The subject manners in this thesis will be discussed in the following order. Firstly, hypotheses in response to the research question will be proposed. These hypotheses are divided between the topics of machine learning and explainable artificial intelligence. Therefore, these subjects will be discussed in separate chapters. Within the respective chapters, theoretical background, method and results are elaborated upon. Lastly, conclusions are drawn and points of discussion, as well as recommendations for future research, are presented.

2 Hypotheses

Theorem 1. *Market share of international book sales within the e-commerce landscape can be predicted with a machine learning model.*

Theorem 2. *The application of XAI techniques can help a business to identify the influencing factors of machine learning models concerning market share in the field of e-commerce.*

3 Market Share Prediction

3.1 Theoretical Framework

Defining Market Share
An incremental step in the research on how certain factors affect the prediction of market share, is to model a way to perform this prediction. Before the predictive modelling can be discussed any further, it is important to clarify exactly what is meant by the variable that is to be predicted. Market share refers to the

percentage of sales in the market made by a specific company. This is shown in the formula below

$$M_{it} = \frac{S_{it}}{\sum S_{it}, S_{jt}, ..., S_{nt}} \times 100$$

where i specifies a firm and t a moment in time [10]. If a company has a high market share this conveys that they hold a stronger position in the market than their competitors. However, the simplicity of the metric and its interpretation makes it futile that the market itself is defined correctly [16]. If the market boundaries are not set in the right way, calculation of market share loses its meaning [13].

Not only is market share a direct reflection of a organization's position in the market, it has also been linked to higher profit. This can be explained because of a high market share loans the company advantages both in terms of economies of scale and market power when entering negotiations with other parties [14].

Market Share Predictors

As stated in the problem statement, traditionally the amount of marketing efforts was seen as the ultimate predictor of market share [24]. However, this is largely due to the focus of traditional market share research being on brands. When taking the market share of a business in consideration, other predictors seem to be more relevant.

A first possible influence on market share is the supply chain logistics of a company. A good logistic relational performance has a positive effect on customer satisfaction, as research into customer loyalty has shown, where an increase of customer satisfaction predicts an increase in customer loyalty. Moreover, that high customer loyalty can lead to higher market share [86]. This indicates that logistics and supply chain can be a predictor of market share, however its direct effect has not yet been measured.

A further possible predictor of market share is price. Price is another factor of which only the indirect effect on market share has been researched. A study has shown that being a competitive company increases the chance of getting a higher market share. Competitiveness is defined as the ability to perform actions to obtain and/or sustain the lead in the market. A price that relatively cheap in comparison to the market, is considered to be a driver of competitiveness [31].

Research has shown that price and logistic performance could predict market share. As market share considers multiple companies, it could prove to be useful to not only use a company's own price and logistic performance to predict the outcome, but also that of important competitors, to the extent of which that data is available.

Machine Learning Taxonomy

Despite being a subfield of Artifical Intelligence already, machine learning is a broad field with a large variety of models that all can be trained in a different way to perform various specified tasks. For this reason, machine learning algorithms

are organized in a set taxonomy, placing algorithms with a similar purpose in the same groups. Firstly, machine learning can be categorized into types. The two main types are supervised and unsupervised learning [4]. Supervised learning makes use of labelled data, where the algorithm aims to map the input to a certain (desired) output. Supervised learning thus focuses on the relationship between the input data and the output. Unsupervised machine learning does not process the labelled part of the data but considers only the input [6].

Furthermore, supervised learning can be organized more extensively based on the task the models are to fulfill. This field knows two primary tasks that it a models aim can be divided into. There are classification and regression tasks. With classification the goal is to map the data into classes that have already been defined, whereas regression models use the input data to map the data to predict a numerical value [67].

The task of market share prediction is one that falls in the field of supervised learning, due to the fact that there is a desired output, market share, as well as a wish to know more about the relationship between the input and this output. Additionally, market share is defined as a numerical value instead of as a category. Therefore, the statement can be made that in order to predict market share a supervised learning regression model is required.

Additional categorization

The consideration of market share is done in connection with the factor of time in order to capture the behavior of the influencing variables [7]. The techniques associated with time series data are therefore useful for the modelling of it. Data falls within the definition of being time-series data when the random data variables are "indexed according to the order they are obtained in time" [83]. The starting point for the model exploration is thus a supervised regression model for time series data.

Models for supervised regression

As mentioned previously, there are many types of machine learning models, even after specifying the task at hand. The purpose of this section is to describe the models explored further in this study, as well as arguing why these models were chosen over other potential models.

Autoregressive Integrated Moving Average Model

An autoregressive integrated moving average (ARIMA) model is regularly used in problems concerning time series prediction. As the name suggests, an ARIMA model combines autoregressive (AR) modelling with the moving average (MA) technique. Through ARIMA modelling the data is explored while flexibly adapting to the structure of the input, which makes it possible to handle different kind of time series [44].

In this model the output is the predicted future value of the input. This predicted future value is assumed to be a function of past observations (AR) in which there is accounted for lagged errors (MA) [101]. ARIMA modelling,

furthermore, requires three parameters: q, d and p. Concerning the AR part of the model, q represents the quantity of data points needed, whereas d signifies the number of steps that it takes for the data to become stationary, and p is the amount of data points needed for the MA part of the ARIMA model [34]. The use of these parameters leads to the model being intrinsically interpretable, the choices that are made in terms of data selection are conveyed through the parameters [98].

Multiple Linear Regression Model

Predictive modelling, in general, is often done through regression analysis. For this kind of analysis multiple linear regression is used in cases where there is on dependent variable and more than one predictor variable [92]. A multiple linear regression estimates the direct effect of a predictor on the dependent variable, while suppressing the indirect effects caused by the other predictors in the model.

The outcome of the model is calculated based on the effects of all the features contingent on their respective weights: a weighted sum [43]. The usage of estimated weights in and for the output of the multiple regression model, makes it that the model loans itself well for interpretability. This is the case because the weights themselves are intrinsically interpretable, needing little to no extra techniques to make a human understand the model's decision-making [88].

Random Forest Regressor

Using a collection of decision tree predictors, a random forest is a supervised learning data mining technique. General model value prediction is done through the average of k of the trees, while also considering the errors of the individual tree predictors and the correlations between these trees [11]. Arguments for the use of random forest base themselves primarily on its performance capabilities. Advantages of a random forest regressor are efficient performance on big training sets, high accuracy even when there is missing data and no overfitting [71]. The randomness of random forests is pronounced in two different ways. Firstly, every tree is based on a random subset of the input data. The second manner is found within each tree, where each further split is based on a random subset of variables [35]. The latter feature of the random forest algorithm therefore means that no tree sees all of the training data, making it easier for the model to avoid overfitting [51].

Even though a random forest is based on a combination of several decision trees, which are a clear example of intrinsically interpretable models, the random forest itself remains a black box. An important reason for this is because of the random ways in which the model composes itself, making it impossible to deduct a clear logic behind its functioning. Besides, there are many possible rules that can be used within the random forest, which are also clouding its interpretability [37]. Consequentially, this does mean that the use of XAI on this model adds a lot of additional value to the original modelling.

Light Gradient Boosting Machine
Another tree-based algorithm is the Gradient Boosting Decision Tree (GBDT), which uses the boosting technique on decision trees. The GBDT algorithm functions in a sequential way, learning from the errors it made previously with the aim of minimizing overall error. This algorithm is especially effective in cases where the data is not very clean [32]. However, the performance of the GBDT model is not suited for large datasets, as the training time is very high. For this reason, the Light Gradient Boosting Machine (LightGBM) was developed, which takes the foundation of GBDT and combines it with other machine learning techniques: Gradient-based One-Side Sampling and Exclusive Feature Bundling. Research has shown that this implementation of tree boosting is more efficient in terms of time, while still upholding the same accuracy standard [49].

As stated, the LightGBM model boosts its ensemble of decision trees to improve the prediction accuracy. This is done by developing the trees with a leaf-wise splitting algorithm, which looks at the subnodes of a tree and focuses on the subnode where there is the most new knowledge to gain from. This is a reason for the speed of the model, however it also comes with the risk of overfitting. Therefore, it is important to control the hyperparameters of the model to prevent this [96].

Comparable to the Random Forest Algorithm, the LightGBM is a black box model, despite it being founded on highly interpretable decision trees. The model does not reveal what factors determine the model's output, nor is the sequence of the decision tree made known. This leads to the model having a low degree of transparency [55]. Nonetheless, this shows that the use of XAI can contribute a lot to the model, which is favorable due to the model's general high accuracy.

Other models
This section will briefly touch upon other supervised regression models fit for time series data and the reasons why these have not been explored further for this thesis.

Artificial Neural Network
Another popular machine learning technique is deep learning, especially in the form of an Artifical Neural Network (ANN). This model consists of multiple layers: one layer with input nodes, one to three hidden layers of nodes and a final layer of output nodes. These layers make it possible to find relations within the input data [95]. Although this model has proven to be highly effective in machine learning problems related to e.g. natural language processing and image recognition, research has shown that tree-based models can out-perform an ANN in other circumstances. Noteworthy examples are datasets with an underlying temporal structure and datasets where every feature holds its own meaning, which both are the case in the current study [57]. Furthermore, Artificial Neural Networks intrinsically have a very low degree of interpretability [72]. Their complexity makes it difficult to interpret

them even using XAI, often needing extra techniques to increase the understanding of the model [50]. Both of these reasons lead to the argument that an ANN model is out of scope for this specific study.

eXtreme Gradient Boosting
Another implementation of GDBT is eXtreme Gradient Boosting (XGBoost). XGBoost is designed in such a way that it can handle sparse data, it can function using parallel learning and the weighted quantile sketch. These additions make it that this model is able to use less processing resources than other existing models [22]. However, in comparison to LightGBM, the other GDBT implementation that has been discussed, the XGBoost has a few disadvantages. XGBoost has been found to have lesser precision, can handle a huge scale of data less well and requires more running time [20]. As this model uses similar techniques to the used LightGBM model, but XGBoost shows some clear disadvantages, it has been decided to not use XGBoosting to predict market share.

Facebook-Prophet
Facebook-Prophet (FB-Prophet) is a model often gravitated towards in the field of retail. In retail the effects of seasonality are often prominent. FB-Prophet is an additive model developed precisely to handle yearly, weekly, monthly and even holiday seasonality effects when considering time series data [102]. For this reason, it would have been favorable to consider this model within the current study. However, due to limitations in terms of data sensitivity and IT access this proved to not be possible.

3.2 Method

In order to predict market share, several machine learning models have been experimented with. These models have been selected based on relevant academic literature. The method section describes the decisions made during the modelling phase and argues that these choices have had a positive impact on the quality of research.

Research Approach
Decisions on data selection, data processing and model evaluation are outlined in these following paragraphs on research approach.

Data selection
The data available for this research has been supplied by the host company. In order to obtain the data concerning the amount of products sold in the market, the host company has contracted a third-party: *Growth from Knowledge* (GFK). The host company provides GFK with their own sales data, and in return receives the sales data combining the entire Dutch book market. This data is aggregated on a weekly basis. The company thus receives the market data of a certain week. In practice, the GFK data has a lag time of two weeks. The GFK data

points used for this thesis are described in Table 1, each specified to the level of European Article Number (EAN).

Feature	**Type**
Number of sales by the market	Numerical
Total revenue of the market	Numerical
Number of sales by the host company	Numerical
Total revenue of the host company	Numerical

Table 1: Externally Collected Data

Based on these values the company's market share is calculated. Market share overall is calculated by dividing all of the host company sales by the sales done by the total market. The same is done for every separate EAN included in the market data for that week. Market share can therefore be looked at on both a market and a product level.

Additional to the data provided by the GFK, the company has provided supplemental internal data. This data is based on the offers of the firm itself, but also on intern data collection of competitors' offers. Within the organization changes are often made daily, leading to the company data differing in value between days, or even within the same date. This data is selected based on the possible predictors indicated by research, as well as recommendations of experts working in the e-commerce field of books. The host company data points used for this study are summed up in the following table.

Feature	**Type**
Selling price of the company	Numerical
Delivery days as shown on site of the company	Numerical
Amount of visits on the product page of the company	Numerical
Selling price of the main competitor	Numerical
Delivery days as shown on the site of the main competitor	Numerical
Release date of a product	Numerical
Company identification code (GID)	Categorical
Language	Categorical
Shop category	Categorical
Valid offer	Boolean
Released	Boolean
Stores closed	Boolean

Table 2: Internally Collected Data

The variables in the internal data table are also each all specified to EAN level. Furthermore, the value of the market share of two weeks ago is added to each EAN in order to simulate the way in which the market data is received. Differences between company and competitor price and delivery days are calculated for every EAN as well, in cases where both of these values are available.

Further included calcuations are the weekly price changes of both the own and main competitor selling prices. The time span covering the data that is selected for this study is all of the 2021 data ranging from 15/02/2021 - 02/01/2022. Earlier data was excluded due to GFK availability issues, later data was excluded as the data selection of this study started in the beginning of 2022. Consequently, any later data did not yet exist at the starting point.

All the data is available on the same EAN level, and through use of this variable they are connected. Nonetheless, there is still a clear discrepancy between the GFK and the internal data. Where the internal data is available on a daily basis, the GFK data is only delivered every week. A choice therefore needed to be made between dividing the GFK data over the week days based on sales division per weekday data or to aggregate the company data over a week. Both of these options would lead to a loss in accuracy for the modified data.

However, in order to combine the data, one of these options still needed to be decided upon. In order to keep the dependent variable of market share intact, the internal data has been aggregated on a week basis. To preserve the product-level specificity, the data has been grouped using a multi-index on the dataframe, which additionally functions as a memory compressor [87]. The dataframe used for this research is thereby indexed with week as the first level and GID as the second level. The numerical values are aggregated by using the average, as an average is able to ignore NaNs and to take into account changing values. Categorical values have been aggregated by using their encoded median, as an average is not representative when a category changes values. Furthermore, in order to reduce memory storage, all categorical values have been label encoded. The processing of the data has resulted in a final dataframe with around 3.2 million rows.

Exploratory Data Analysis

For the purpose of exploring and evaluating the data, an exploratory data analysis (EDA) has been conducted. The EDA can help provide information on possible drivers of market share. Furthermore, EDA can be insightful by showing correlations between predictors as well as between the independent and dependent variable [73]. These correlations have been looked at through the use of the Pearson correlation coefficient, where a correlation is considered to be strong when the coefficient value is above 0.7 [81]. Additionally, during the exploratory data analysis phase the market share data is tested for seasonality. With the use of time series data, it is important to check whether this data is influenced by seasonality, to guarantee that models are selected

that fit the pattern of the data [70]. This test is conducted through the seasonal decompose function from the python statsmodels package [82].

Validation
Model validation is achieved best by splitting the data into train and test set(s). In machine learning cases where the observed data is independent and distributed evenly, cross-validation is a common method to achieve this [80]. However, in cases where time series data is used, these two requirements do not hold up. Therefore, another method is needed. Other studies interested in machine learning prediction in the field of retail have chosen for the rolling window approach, especially because of its ability to capture the dynamics of price change [21,97]. With the rolling window approach, a window of size $0 < W \leq N$ is used to constantly recapture the training and test data. The specified window is used to predict the data one time series step after the window [54]. In the case of this research that step is a week. Besides the previous advantages mentioned, another benefit of rolling window usage is that the model constantly updates. Within these updates data that is older and less relevant is neglected, thereby focusing on the data closest to the point of prediction [90]. Based on experiments during the modelling phase, the window size was set on 4 weeks, which equates to around one month. Another model validation method could be k-fold splitting, where the data gets split into k amount of folds. However, without the possibility of performing k-fold cross validation splitting, which is due to the nature of time series, this method of splitting has an increased risk of both over-training and lack of generalization [78].

Accuracy
In order to compare between the performance of the selected models an accuracy measure had to be chosen. The main accuracy measure decided upon for this study is that of Root Mean Square Error (RMSE). The RMSE can be defined as the standard sample deviation between the actual and the predicted error, and is calculated with the following formula:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Predicted_i - Actual_i)^2}{n}}$$

The closer to 0 the RMSE is, the more accurate the model predictions are. An advantage of this measure is that RMSE is in the same unit as the output variable, making it easier to understand how the error is distributed [60]. An argument stated against the use of RMSE is that this measure is less capable of dealing with outliers than other accuracy measures. This argument is relevant to take into account, due to the fact that the book market includes a very big long tail, leading to a big group of outliers. However, research has proven that in cases where there are $n \geq 100$ data points, this argument no longer holds, as is the case in the current study [19]. The accuracy measure R-squared will

be used as a secondary, control accuracy measure. The R-squared value ranges between 0 and 1, serving as a percentage of the squared correlation between the predicted and actual values [60].

Machine Learning Models

Autoregressive Integrated Moving Average

In order to fit the research data to the ARIMA model, the input data was transformed to a dataframe indexed on Week and with market share as only column. The ARIMA model thus looks at the prediction of market share overall. Using the Augmented Dickey-Fuller test [85], it is concluded that this dataset is stationary ($p < 0.05$). Consequentially, no steps are needed to become stationary so $d = 0$. By studying a Partial Autocorrelation and a Autocorrelation plot, the parameters $p = 1$ and $q = 1$ are determined. Therefore, the parameter order of the ARIMA model is (1,0,1). In order to make an ARIMA prediction, the python package 'statsmodels' has been used [82].

Multiple Linear Regression

A multiple linear regression is able to predict market share on product level. The data is therefore structured with a Week-GID multi-index, as described previously. However, the structured data still required extra pre-processing before being suitable for the model. Using the python sci-kit preprocessing package, all of the numerical columns were normalized using a standard scaler [9]. Furthermore, the non-numerical columns were transformed into dummy variables using a one-hot encoder, because a linear regression is not able to work with categorical data [84]. Lastly, a limit of the multiple linear regression model is that it is not able to handle missing (NaN) values [93]. For this reason, all of the rows with one or more missing values were dropped from the input, resulting in a dataframe of nearly 2 million rows. The model execution is done by using sci-kit's linear regression function [9].

Random Forest Regressor

A random forest regressor model holds some similarities to a multiple linear regression. Market share can also be predicted on a product level by this model, therefore holding the same Week-GID multi-index structure. Likewise, a random forest is unable to process categorical values simultaneously with numerical values [11]. For this reason these columns were transformed through one-hot encoding. Although random forest are capable of dealing with missing values in theory [42], practical implications can often not deal with this. Missing values were dropped from the data as a consequence. As a last step all numerical columns were scaled by use of a standard scaler. The random forest model has been implemented in python with the sci-kit ensemble package [8].

Light Gradient Boosting Machine

The LightGBM model is capable to predict market share on product level, allowing the data to be structured with a Week-GID multi-index. As explained

earlier in the theoretical framework, LightGBM is a version of a Gradient Boosting Decision Tree extended with Gradient-based One-Side Sampling and Exclusive Feature Bundling. The latter makes it possible for the model to internally handle categorical data, without need for external encoding. Research has shown this internal processing of categorical values works significantly better for a LightGBM model than the external handling [41]. Therefore, the same has been done in this study. Furthermore, the LightGBM also has a default manner of dealing with missing values. All of the cases having a value missing are replaced by NaN values, which the LightGBM model is able to handle. For this reason no further processing has been employed on the missing values. The Python LightGBM package was used to implement the model [63]. Lastly, the LightGBM functions with hyper-parameters, whose tuning can help prevent overfitting [96]. To achieve this tuning the Python Package Optuna was used, which is a software developed for the parameter tuning process [2]. The hyper-parameter tuning resulted in the parameters shown in the table below.

Hyper-parameter	**Result**
metric	RMSE
boosting type	GBDT
learning_rate	0.03
num_leaves	52
n_estimators	1596
colsample_bytree	0.5
min_child_samples	32

Table 3: LightGBM Hyper-Parameters

The *metric* and *boosting type* hyper-parameters were determined before the tuning. These represent the metric used to measure accuracy, and the GBDT on which the LightGBM model was based. The hyper-parameter *n_leaves* is the number of leaves used per tree, which in the case of the current study are 52. The *learning rat*e controls the speed of model iterating [89]. How many decision trees are combined is determined by the parameter *n_estimators*. The number of *colsample_bytree* indicates the percentage of features that are used for the building of a singular tree [99]. The last hyper-parameter, *min_child_samples*, is an indication of the minimum amount of data needed in a leaf, which is set to 32 for the model used in this study[65].

Research Quality

As stated in the subsection defining market share, it is crucial to cover the correct market when calculating market share. If this is not done right, the metric loses its meaning. The GFK guarantees that their collected data covers parties that

together account for at least 90% of the total Dutch book market sales. The missing parties are primarily local book stores. As this study focuses on the online Dutch book market only, the assumption can be made that the data has a coverage of at least this 90%. Therefore, the quality of the data coverage is seen as sufficient to use for the market share prediction.

However, while most of the market sales are covered, the same can not be said for all of the market events. The collected data is limited to the variables described in the subsection data selection. Therefore, it can be that unforeseen variables are not included in the data sets. The possibility thus exists that events have occurred that did have an effect on market share, but can not be taken into account during the predictive modelling.

3.3 Results

Exploratory Data Analysis

The first step in the Exploratory Data Analysis was to look at the course over the year of the company's overall market share. This is illustrated in figure 2. In this figure it is shown that the company's market share in general ranges between 70 and 80%. There are however, some drops in the value midway through the year and at the end of the year.

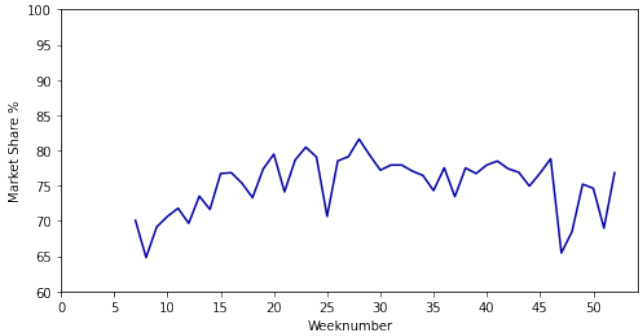


Fig. 2: Company Market Share in 2021

These drops in market share are also reflected when considering the sales in the market in comparison to the company's sales, as seen in figure 3[3]. In this graph a lot of similarities are seen between the market and the company, which can be explained by the fact that the company is a large player in said market. Although the company and the market sales do seem to follow the same trend, in the weeks where the company is lacking market share the market seems to either shrink less or grow more than the company.

[3] Due to confidentiality reasons the y-axis has been hidden from the display of the graph.

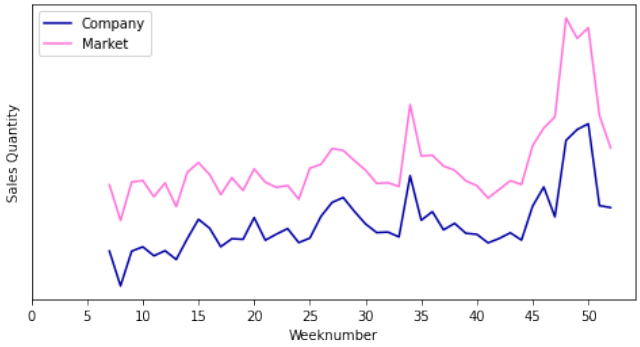


Fig. 3: Market vs. Company Sales

As stated, literature on company market share indicates that delivery and price could be factors that are able to affect market share. For this reason, in the EDA different comparative buckets between the company and its main competitor have been made for both of these factors. How these buckets perform in terms of market share in comparison to the company's total market share can be seen in figures 4 and 5.

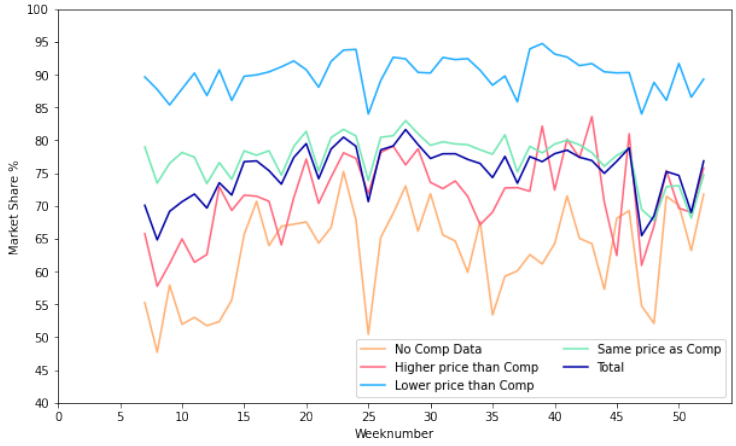


Fig. 4: Selling Price Comparison

Figure 4 shows the market share the company has over the respective price buckets. Remarkable observations from this graph are especially the buckets (i) lower price than competitor and (ii) no competitor data. Looking at the market share for books that the company sells cheaper than the main competitor, it is clear that this is much higher than the company's total book market share.

This indicates that a relatively low prices can influence market share. On the other side, the market share of the bucket where competitor data is unavailable is much lower than that of the company. This could indicate the importance of knowledge the behavior of other players in the market.

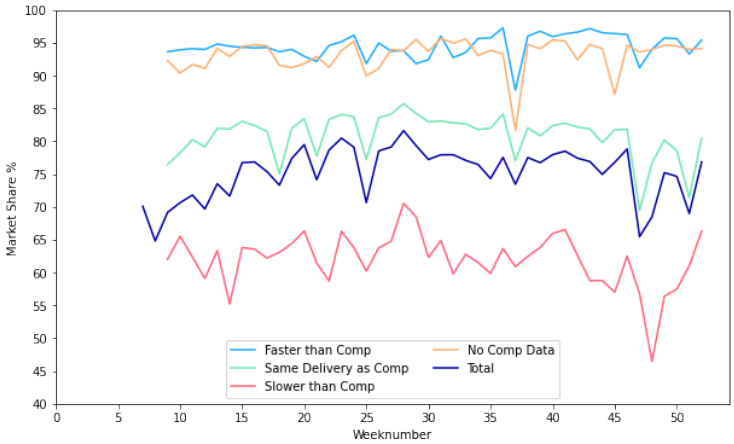


Fig. 5: Delivery Time Comparison

However, figure 5 shows that the bucket covering no knowledge on the competitor's delivery time does not have the same low market share as no data on the competitor's selling price does. Furthermore, the bucket of the company delivering faster than its competitor seems to have a higher market share than the bucket of the competitor being cheaper than the competitor does. Similarly, when a product has a slower delivery time than the competitor, there seems to be a lower market share than when a product has a higher price than that of the competitor. Both of these observations indicate that selling price comparison and delivery time comparison have different dynamics.

The EDA additionally covered variable correlation and distribution. No strong Pearson correlations were found among the predictor variables nor between the predictor variables and the dependent variable. Further EDA results concerning the distribution of values can be observed in Appendix A.1: Extended Exploratory Data Analysis.

Outliers

Throughout the data analysis, especially while looking at the data distribution, some outliers surfaced. After a check on the impact of these outliers on the market's total revenue, which was 3 percent, the outliers have been removed from the modelling set. All products that were sold less than four times per year or more than a thousand times have been excluded, as they were not representative

for the market share dynamic and could cloud the model results. The included products in total represent 97% of the book market revenue.

Seasonality
At last, the EDA tested whether there was any seasonality in the provided data. The results of the seasonality decomposition test can be found in the figure below. In this figure it is shown that the seasonality line is flat, indicating that no seasonal effects were found in the study's dataset. However, the test could only look at weekly or monthly seasonality. Due to the limited timeframe of included data, only a singular year was included. This makes it impossible to verify whether there was any yearly seasonality effect at play. Therefore, it can not be stated that seasonality was not of influence on the prediction, as the data only includes a limited set of weeks. Nonetheless, no seasonality was found within the current set, therefore it will not be taken into accounted for the predicting models.

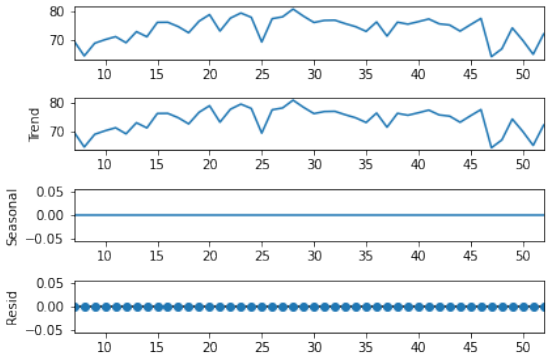


Fig. 6: Seasonality Decomposition

Machine Learning Model Results
In the following sections the results of the different machine learning models will be discussed. These results will be discussed in terms of the way the models processed the data and based on their accuracy. Furthermore, a few plots will be used in consideration of the model performance. The first plot used for the model results is a predicted versus actuals plot. In these plots the actual values and the predicted values are both plotted on other sides of a diagonal line. When looking at the plot, its shape can provide information on how well the prediction is [30]. A second plot that will be used to evaluate most models is one showing the residuals distribution. In this graph the residuals, which represent the distances between the actual and the predicted values, are plotted against the actual values. Furthermore the distribution of these residual values are displayed [3].

Autoregressive Integrated Moving Average Model

The results of predicting market share with the ARIMA model are shown in figure 7. As seen in this graph, the ARIMA forecast often differs from the actual values in the test set, especially halfway (week 26) and towards the end of the year (week 47). This corresponds to the market share differing from the general trend, as discussed in the EDA. The RMSE of the ARIMA model is 4.62, which can be considered to be a fairly good score as the market share values can range from 0-100. However, for the ARIMA model, the market share input range was not extended to this same scale. This is explained by the fact that this model has looked at market share overall, only using a (64.23 - 80.82) range. Taking this in consideration, a 4.62 on a scale from 0-16 is by far not as good of a score.

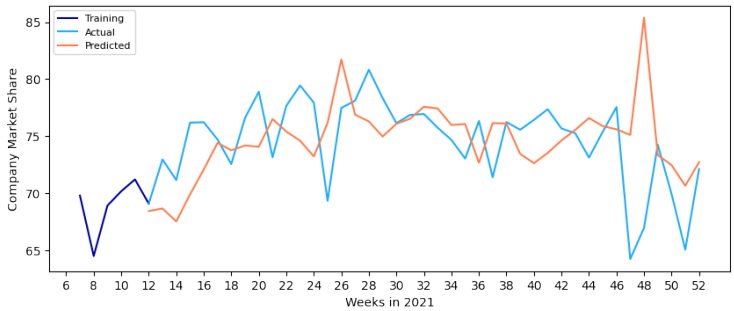


Fig. 7: ARIMA Prediction Output

Furthermore, as the ARIMA model considers only the market share overall, there is not a lot of input data. This is also reflected in the predicted versus actual values plot seen in figure 8, where the amount of data points are limited. Nevertheless, figure 8 is still able to show us that the ARIMA prediction lacks in accuracy. Instead of the predicted and actual values mirroring each other, the points appear to be scattered in a shape that more resembles a square.

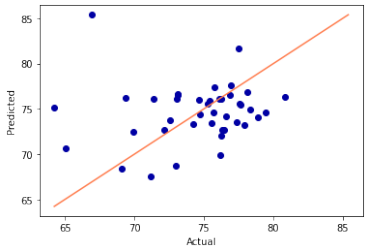


Fig. 8: ARIMA Predicted vs. Actual Plot

To summarize, the results of the ARIMA model include a high RMSE score and a distorted predicted versus actual plot. This shows that the ARIMA model is not a suitable model to predict market share in this study. This can be explained by the limited amount of predictors used by this model, namely only previous market share. Literature and field knowledge have indicated there are other factors that should also have an influence, which is confirmed by these results. Past market share alone is not able to predict future market share. Furthermore, a limitation of the ARIMA model is that the predictions were made on market share overall. Consequentially, the model cannot be split to look at a specific group of products and their course of market share.

Multiple Linear Regression

A model that is able to predict market share on a product level is the multiple linear regression. This is reflected clearly in figure 9, where there are a multitude of data points. However, what also becomes clear when considering this figure, is that the multiple linear regression is not able to predict this product level market share sufficiently. Firstly, this is shown by the relative large amount of predicted values higher than 100. For an actual value it is not possible to have more than 100% market share, which is not reflected in the results produced by the model. Secondly, the diagonal line helps show that the predicted values do not mimic the actual values, especially with the above and below average market shares.

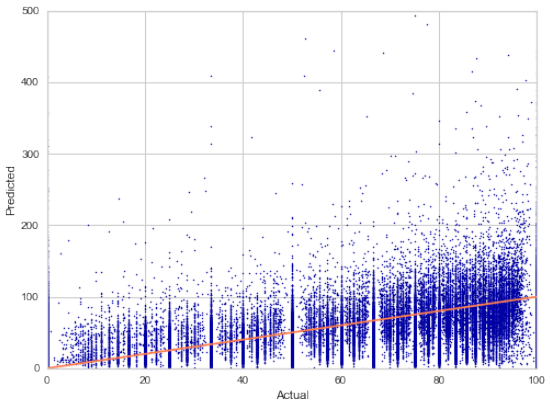


Fig. 9: Multiple Linear Regression Predicted vs. Actual Plot

Corresponding to the figure above, the RMSE of the total Multi-Linear Regression model is 32.87, on a scale from 0-100. This metric shows that the accuracy of the model is not high, as there is quite a large error still. An explanation for this error can be found in the residuals graph in figure 10. The residuals of the model are not plotted around the 0 line, but they are shaped in

a diagonal form over the graph. Moreover, similar to figure 10 it is shown that the predicted points go much higher than the maximum possible larger market share, with even residuals above 200. These divergent residuals indicate that the model is unable to convert its input to an accurate prediction.

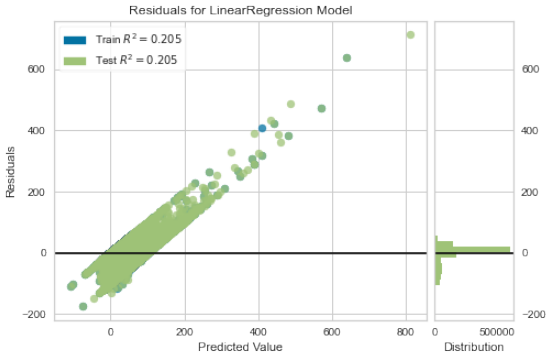


Fig. 10: Multiple Linear Regression Residuals Plot

Even though a multiple linear regression is capable of predicting market share on a product level, this results section has shown that the model is not able to do this prediction accurately. The model results in a high RMSE score and a misshaped predicted versus actuals plot. Moreover, the model's results did not correspond with the fact that market share can not be higher than 100%. Therefore, it proceeded to predict many values above this threshold.

Random Forest

The Random Forest Regression is able to predict market share on a product level. Moreover, a random forest is able to learn the range of values that are possible for market share. This is reflected in the figure below, where the predicted points do not go above the value of 100 nor below the value of 0. However, the predictions within this range still appear to lack in accuracy. Especially when considering the predictions where the company has a market share value of lower than 70%. This is reflected both in the shape of the predicted versus actual points in Figure 11, as well as in the total RMSE of 30.10, averaging over all of the predictions.

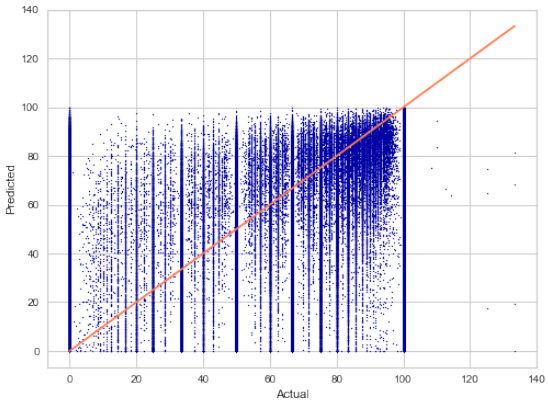


Fig. 11: Random Forest Regression Predicted vs. Actual Plot

Diving deeper into the model's results is possible with consideration of the residuals plot displayed in the figure below. In this figure the distribution of residuals can be found, illustrating that these are centered around 0. Furthermore, the distribution is skewed to the left, which corresponds to positive residual values. This shows that the random forest model often predicts a market share value as too low rather than as too high.

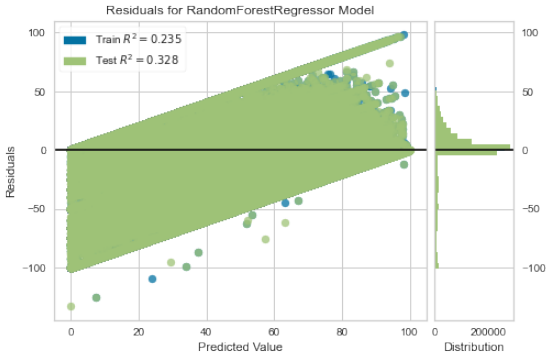


Fig. 12: Random Forest Regression Residuals Plot

A random forest algorithm has the ability to predict market share on a product level. Furthermore, the model is able to pick up the range in which the values of this variable should be predicted. Despite this learning ability, the model's predictions are still lacking in accuracy. This is especially reflected by the model under-estimating the company's market share. Lastly, due to model-specific limitations, not all products could be included in the model.

Light Gradient Boosting Machine

The amount of values in figure 13 in combination with the range they fall into shows that a Light Gradient Boosting Machine is able to predict market share on a product level within an appropriate range. Nevertheless, some predicted values still appear to be above the maximum market share percentage of 100%. The majority, however, remains to be in the defined range. Furthermore, unlike with other models implemented in this study, the LightGBM was able to process all of the available data, even when in rows some values were missing. This could be a possible explanation for the anomalities. Despite the increased amount of data included in the model's input the RMSE is 25.15, which can be considered reasonable. Additionally, when considering the graph below, the values seem to mirror each other in the cases where the company has more than half of the market share. Below this value, the plot shows that less of a match.

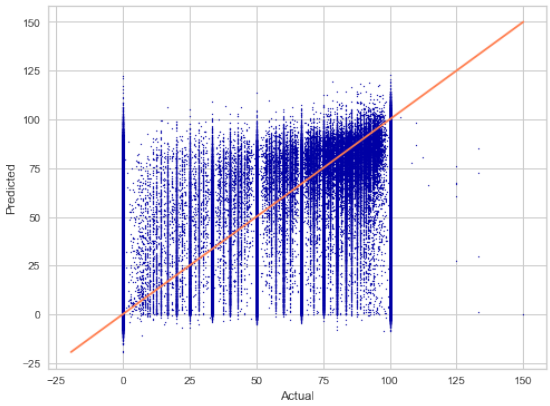


Fig. 13: LightGBM Predicted vs. Actual Plot

Further exploration of the residuals of the model's outcome confirms that the model predicts close to the market share range, although it does show the deviations the predictions have. The residual distribution demonstrates a skew towards the positive values on the left of the 0-axis. This skew shows that the model often predicts a value as too low, and barely predicts market share values as too high. Furthermore, it is shown that while there is this skew to the left, there is also a long tail of distribution running to the right. This is different than the left side of the distribution, where the residuals appear to not go over the value of 50.

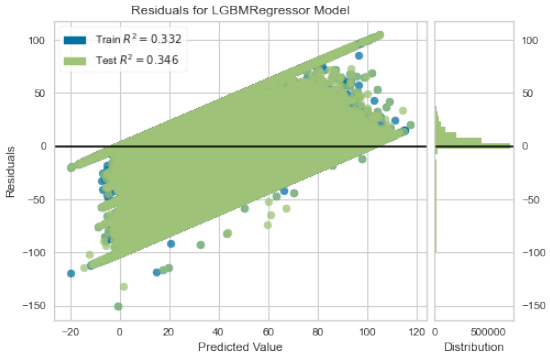


Fig. 14: LightGBM Residuals Plot

The LightGBM model has shown promising results, both when considering the scope of data it can predict as well as in terms of accuracy. However, the literature on the LightGBM model praised it for its very high accuracy, which is not immediately reflected in the model's results. A further exploration of the course of the overall RMSE shows is portrayed in figure 15. This figure shows a clear and high deviation in week 37, as well as weeks 51 and 52.

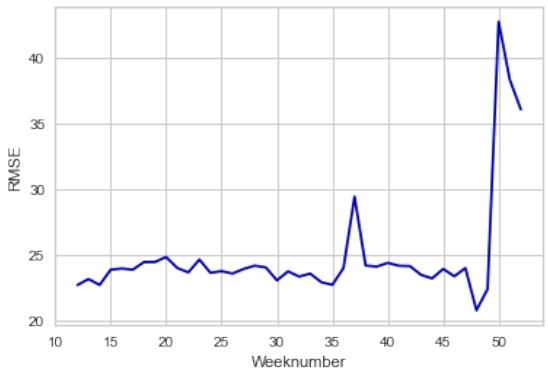
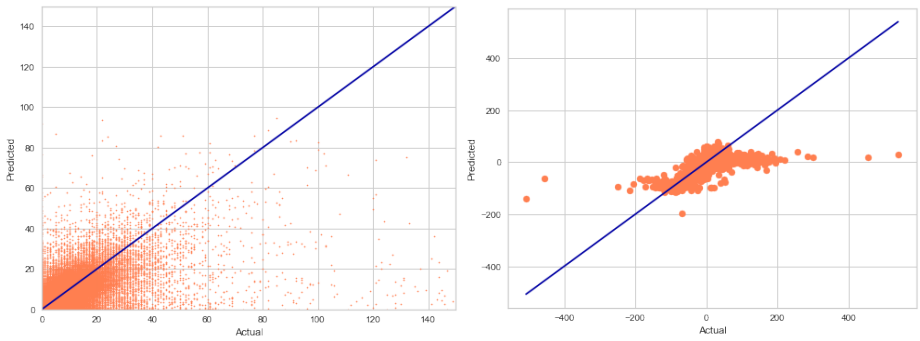


Fig. 15: LightGBM RMSE Course

Further exploration of the mentioned weeks led to understanding of the resulting deviations. In these weeks the host company's most important competitor held promotion weeks. This promotion campaign could have resulted in more traffic to the competitor's website, better placing of the competitor in the market and other positive consequences for this competitor. Together this could have led to a decrease in company market share, unexpected by the model.

The impact of these weeks was assessed by re-running the LightGBM model function. However, in this iteration the deviating weeks were dropped from the data used by the model. Without having to process or predict these weeks the model's RMSE results to 23.64, which is less than the RMSE of the LightGBM model that includes all weeks. As the only difference between these models are whether they include the described weeks, this can be seen as the impact that the competitor's promotion weeks have on the LightGBM model's accuracy.

The inability of the model to comprehend competitor events in certain weeks partially explains the higher than expected market share. Besides this, a possible additional explanation can lie in the nature of the predicted variable. To know market share, is not only to know your own sales, but also know how they compare to the rest of the market. A challenge in the prediction is the limited amount of competitor data in comparison to own data. To inspect whether this limitation is an explanation for the model's accuracy score, the company sales, as well as the shift in sales have been predicted.



Company Sales Shift in Company Sales

Fig. 16: LightGBM Sales Predicted vs. Actual Plots

Figure 16 showcases the predicted versus actual plots of the sales and shift in sales predictions. These plots reflect that with the input data, a LightGBM model is able to predict the company sales sufficiently (RMSE is 1.99[4]). The figure shows that predicted values mirror the actual sales. When the model aims to predict the change in company sales, the mirroring is already a bit less clear. As shown in the plot on the right, the shift in sales is mirrored well between the values in the range (-100, 100). However, this mirroring seems to fade when there has been a large shift in sales. The RMSE of shift in sales is 2.05 [5]. These results support the argument that the model's predictions are impaired by the limited amount of competitor data available, as the model itself is accurate in predicting the 'own sales' factor of market share.

[4] On a range from 0 to 1000.
[5] On a range from -1000 to 1000

To summarize, the LightGBM model shows that it can predict market share on a product level, while mostly understanding the numerical range in which market share can exist. Furthermore, the model is able to make these predictions using the entire data set, as it holds no limitations for missing values. Although the model results have an acceptable accuracy, the RMSE was relatively high in comparison to the expectations set by previous literature. Possible explorations for this were explored. A first discovery showed that certain weeks hold a significantly higher RMSE than the general RMSE course, due to competitor promotions that are unknown to the data collected by the business and therefore the model. Secondly, a limitation for the prediction is that competitor data is not available to the same extent that own data is, while both are needed for the market share calcuation. The model is able to predict own sales quite accurately, but does not have all the necessary data to detect changes in the market.

3.4 Machine Learning Conclusion

The results have shown that the selected models are all capable of predicting market share to a certain extent. However, the models differ in terms of approach, their handling of data, and the accuracy of their predictions. In comparison to the others, the ARIMA Model had a low performance in these categories. The model was unable to handle the data on a product level, and therefore had a RMSE that was difficult to compare to the models that were able to make a product-level based prediction. Although the Multiple Linear Regression model was able to make a prediction on a product level, it did not perform well when considering its accuracy. The predicted versus actuals plot appeared distorted, due to the predicted market share range, and the RMSE was not satisfactory. The accuracy of the Random Forest Regressor performed a bit better, with especially the predicted versus actuals plot forming a more sensible shape due to a correct range. However the RMSE was still above 30. Furthermore, neither of the models discussed so far had the adequacy to process all of the data available, and therefore were unable to make weekly predictions of every product in the dataset. This is not true for the Light Gradient Boosting Machine, which is able to handle missing values. Despite working with partially missing data, this model was able to perform best in terms of accuracy. A more detailed look into the model's results showed that the accuracy was raised by three specific weeks. This could mean that in these weeks certain events occurred that have not been foreseen by the model or the business. An additional finding was that the model was very capable of predicting the company's own sales. Consequently, the market sales part of market share seem to be the model's weak point. This can be explained by the limited amount of competitor data available, especially in comparison to what company data is available.

Based on the model comparison, the LightGBM model has been selected to further explore for Explainable Artificial Intelligence. The model outperforms the other models when considering both accuracy and completeness of the predictions. Furthermore, an analysis of the model's results have provided the

confidence that the model is able to predict the majority of the market share data sufficiently for the goals of this study. Lastly, despite the results analysis, the model remains a black-box, which makes it interesting to discover what features were influential for the model's predictive modelling.

4 Explainable Artificial Intelligence

The current study has resulted in the development of a model capable of market share prediction on a product level, using all the available data. Nonetheless, the functioning of this model remains a black box, it is unclear how and why the model came to this prediction. Consequently, this means that no conclusions can be drawn on the importance of the separate market share predictors. For a business to be able understand the model's decision-making, Explainable Artificial Intelligence is introduced.

4.1 Theoretical Framework

Defining Explainable Artificial Intelligence

Explainable Artifical Intelligence (XAI) can be described as a field aiming to make other forms of Artificial Intelligence more understandable to humans [1]. As explained in the problem statement, this increase of understanding can help solve the black-box problem that often occurs when using machine learning, making developed models more transparent. In describing XAI, many terms are used to describe whether a human understands a machine learning model, which can cause confusion [12]. In order to prevent sortlike confusion within the current study, a table defining these terms is provided below.

Term	**Definition**
Understandability	A human can comprehend the model without requiring to know more about the algorithmic background of this model.
Interpretability	The ability to explain the results of a model in a way that a human can comprehend.
Explainability	The explanation of a model that represents how decisions within the model were made in a way that humans can comprehend.

Table 4: XAI Terminology [5]

Explainable Artificial Intelligence focuses on creating explainability, the XAI techniques aim to explain the model's decision-making in such a way that the audience of the explanation now comprehends how the model came to these decisions. It is important to note that this human comprehension is central in whether or not an XAI explanation is successful [56].

Explainable Artificial Intelligence Taxonomy

Similarly to machine learning, the field XAI aims to help comprehend, XAI can occur in many forms. In order to select the right type of XAI methods for the explanation of a machine learning model, it is important to understand what categories of these methods exist and how they correspond to developed models. These will be highlighted in the following section, focusing on which methods fit the chosen machine learning model.

As touched upon briefly in the problem statement, the first classification of XAI methods considers in which part of the modelling phase they were implemented. The three phases during which this is possible are the pre-model phase, the in-model phase and post-model. Methods implemented during the pre-model phase are those that are independent of the chosen model, as they focus on the available data only. If the model is explainable during the in-model phase, this is due to the model being intrinsically interpretable. These models are transparent in such a way that no additional techniques are required to understand how the model works. Lastly, there is the post-model phase (also referred to as post-hoc interpretability), which aims to better a model's interpretability after it has been built [17]. The current study focuses on a LightGBM model as the previous section showed it had the best performance out of the models experimented with. Therefore the pre-modelling and the in-modelling methods are excluded from the scope: we want to look at explainability dependent on a model and this LightGBM model is not intrinsically interpretable.

A further, related, classification of interpretability methods is whether a method is model-specific or model-agnostic. The criteria for a model-agnostic method is that it should be applicable on any model, whereas a model-specific method applies to one model only. This often results in model-specific methods being used during the in-model phase and the model-agnostic methods used post-hoc [66]. The current study looks at the LightGBM model, which needs post-hoc explainability and therefore a model-agnostic method matches this model best.

Thirdly, a XAI method can increase the interpretability of a model on either a global or a local level. Global interpretability refers to whether it is possible for a human to understand how the model produced its results in general. For a method to provide local interpretability it must illustrate how a model came to an individual prediction and why it made that decision [28]. Within the context of this study both of these types of methods can provide interesting results, therefore both will be considered for usage.

The last classification between methods based on the type of results a method offers. Literature divides these results in four categories, as expanded on in the list below [17,66].

1. Model internals. These results are those that are provided by intrinsically interpretable models. By definition all of these results are model-specific.
2. Data points. Within these categories fall all models that result into a data point that increases model interpretability. This data point can be either

new or already existing. These type of methods work in cases where these data points themselves are interpretable. Often, for tabular data this is not possible and as a consequence methods that fall into this are primarily used for image or text processing models.

3. Surrogate intrinsic interpretability. Methods that fit this category are those that simulate intrinsically interpretable models to estimate a model which lacks this in-model interpretability. This estimation can further provide explanations, either global or local, on this black box model.
4. Feature summary. When a method returns a summarizing statistic of features in the model, it falls under this last category. Results that fall under this category can either be numerical values, such as feature importance, or can be a summary that is only sensible after visualization. An example of the latter are partial dependence plots.

Taking the model that has been selected for further explanation into account, the first two categories do not seem appropriate to use. The LightGBM model that predicts market share is not an intrinsically interpretable model, and is built using tabular data only. For these reasons those result categories are considered out of scope for this study. In consideration of the goal of the current study, the last result category seems most important, for this matches best with gaining knowledge on market share on influencing factors. Therefore the study will focus on this aspect of the XAI methods. Nonetheless, the surrogate intrinsic interpretable results have to potential to provide a different perspective in terms of gaining this knowledge. For this reason, this category is also included.

Explainable Artificial Intelligence Audience

Even though the goal of Explainable AI is to increase the understanding of a complex model to its audience, this goal is not always reached. In the business context the stakeholders that are making decisions based on the models often do not have a technical background. The explanations made by the XAI methods can prove to be too complex for this non-technical audience, thereby failing to improve the interpretability of the model. Jiang and Senge describe this phenomenon as a gap between two 'XAI cultures', what might be enough understanding for a data scientist who developed the XAI might not be a sufficient explanation for a business stakeholder [47]. This gap can prove to be an issue when employing XAI as a driver of business strategy.

Not only does the background of the intended audience matter, so do the needs that this audience has. A meta-analysis on XAI audience goals has defined the following possible needs: trustworthiness, causality, transferability, informativeness, confidence, fairness, accessibility, interactivity and privacy awareness [5]. Out of these goals, fairness and privacy awareness are not relevant for the current study as they are the goals of regulatory agencies. Considering the other needs, the one that appears closest related to the formulated company goal is informativeness. Informativeness is the need to have an overview of available information that is able to support

decision-making. XAI is able to provide an overview of this information, so that its audience understands how the model acts and is able to make decisions based on this information [45]. As stated, the company indeed requires information that can support the decisions they need to make to reach market share increase.

In their study on XAI methodology, Vermeire et al. [94] highlight the need of matching your stakeholders to the explainability method closest related to their needs. However, no clear method is defined on the process of this mapping. Methods that have been chosen for this study, however, are those that (a) fit the selected LightGBM model and (b) can steer towards the defined audience goal of informativeness. These models all provide different types of information, as there has not yet been a previous study showing in what way business stakeholders would like to be informed.

XAI Methods for Post-Hoc Interpretability

Feature importance
Feature importance is used to describe how important a certain feature is for the predictions performed by a certain model, without accounting for whether the model is linear or what the direction of the importance is [18]. An advantage of looking at a model's features in this simple way, is that it is easy to implement as well as easy to understand [96]. On the other hand, a disadvantage of this traditional way of looking at the features of a model, is that it can only provide information on their respective importance. It can not show how exactly a feature impacts the model [53]. This could make it difficult to use the information to support the decision-making process, as the information itself is quite limited.

SHapley Additive exPlanations Framework
An Explainable AI method that stems from game theory is that of SHapley Additive exPlanations (SHAP). With this method, every feature of the model is given a SHAP value. A SHAP value is "the average marginal contribution of a feature value across all possible coalitions" [66]. This represents the extent to which the model's prediction would change when the model would be conditioned on that feature. SHAP is a model-agnostic method, which can be used both in a global as well as a local manner. The desired result in either case is that of feature importance [58]. The fact that the values for every feature are calculated in the same way makes it possible to order them based on importance. Furthermore, the calculated results can help expose patterns within the data. Not only is it possible to look at individual features through use of SHAP, it is also an option to look at their interaction effects. These interaction effects can be defined as the combined feature effect after taking their individual effects into account [66]. This makes it possible to learn more about the ways through which the features are able to impact the model's outcome.

An argument in favour of the use of SHAP is that it has a mathematical background, which helps the understanding of the method itself [61]. Moreover,

Lundberg et al. show in their study, focused on interpretability for tree-based models, that SHAP visualizations are intuitively easy for humans to understand [57]. This could prove to be an asset in the current study, as the XAI audience is one with a non-technical background.

Local Interpretable Model-Agnostic Explanations Framework
The Local Interpretable Model-Agnostic Explanations Framework (LIME) uses its name to reveal in which XAI classification it belongs. It is a local, model-agnostic model which offers the results as a surrogacy intrinsicically interpretable way. The goal of LIME is to provide an interpretable explanation on a local level, the surrogate models are thus trained to explain individual predictions [79]. For tabular data, the framework describes the importance of the features for the predictive task. To achieve this, LIME does not make use of the model's training set, but rather develops its own training set by taking alternative combinations from the data and analysing how every combination would influence the prediction [66].

Dieber and Kerrane have studied the framework's strength and limitations. Advantages of LIME usage appear to be its clear increase in interpretability among its users, even when they have no prior XAI knowledge. However, there also appear to be some disadvantages. Despite the interpretability increase, there are no clear guidelines on how to understand the LIME output. This can cause frustration among its users. Furthermore, due to the local nature of the framework, it is hard to make any global statements about the model. If one wanted to achieve this, additional manual work would be required [26].

Decision tree
Originally, a decision tree is a supervised machine learning model itself. This model works without parameters, but instead uses an top-down approach focusing on the chances of possible outcomes. As a result, the model's output consists out of conditional decision-making rules [59]. These straightforward rules on how the model makes its predictions are thus part of the model's output. For this reason, a decision tree has a very high degree of interpretability [38]. This is also illustrated in the figure below, where several machine learning models are plotted on interpretability and accuracy axes. In this graph it becomes clear that a decision tree might not have a high degree of accuracy in terms of prediction, it is already interpretable on itself and does not require any post-hoc XAI methods. The low degree of accuracy can be explained by that a decision tree is inclined to overfit [75].

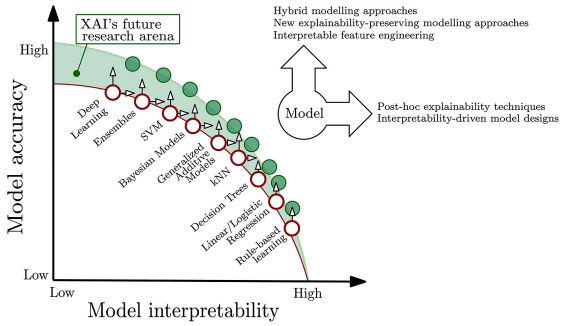


Fig. 17: Interpretability vs. Accuracy Comparison [72]

However, a decision tree has also been proven to be one of those post-hoc XAI methods itself. When done so, a decision tree becomes a post-hoc, global XAI method presenting surrogate intrinsic interpretability. The output of this method is often an answer to how the model functions and whether it appears to be generalizable [25]. An advantage of employing this method is the straightforward nature of the output, making it easy to understand for both technical and non-technical stakeholders [39]. Nonetheless, there are also some disadvantages. A first disadvantage of using a decision tree as an interpretable approximation of a black-box model, is that the interpretability of the decision tree is limited. The more depth within the decision tree, the less interpretable it becomes. This rule applies to any audience interpreting the tree, regardless of their background [46]. Furthermore, as the input of the decision tree is another model, it is not possible to draw conclusion on the data used for the original black-box model. Lastly, there are no clear guidelines on when a global surrogate model, like a decision tree, approximates the original model good enough. This can make it difficult to state clear confidence in the interpretations made [66].

4.2 Method

In order to explain the LightGBM model, several Explainable Artifical Intelligence methods have been employed. These methods have been selected based on relevant academic literature, as expanded upon previously. Furthermore, qualitative methods have been used in order to evaluate the audience's response to these methods of XAI. The method section describes the decisions made during this phase of the study and discusses the impact of these choices on the quality of research.

Research Approach

Decisions on XAI framework implementation and interpretability evaluation are outlined in the following paragraphs on research approach.

Interpretability Frameworks

Feature Importance

Computing feature importance in a traditional manner is a straigh-forward process. The LightGBM is fit to the training data, and can then be used as input. The python lightgbm package includes a plot importance function that makes it possible to immediately plot the feature importance with the fitted model [64].

SHapley Additive exPlanations Framework

In order to implement the SHAP framework two different calculations needed to be made. Firstly, the SHAP values of every feature needed to be calculated, and secondly the SHAP value of every feature interaction was calculated as well. The feature SHAP values were used to create a feature summary plot. Dependency plots of interesting interactions were created based on the SHAP interaction values. The input for the SHAP explainer is the LightGBM model fit to the training data. Although SHAP is able to provide both local and global explanations, the main focus within this study will lie upon its global functions. This is in order to provide a contrast to other local methods. To compute the SHAP values the python shap package was employed. An advantage of this package is that it is highly compatible with LightGBM models [48].

Local Interpretable Model-Agnostic Explanations Framework

Before the LIME framework could be implemented, some pre-processing was required. Firstly, LIME is incapable of dealing with rows that have missing values [40]. For this reason those rows have been filtered out from the dataset before the LightGBM model was trained. A separate version of this LightGBM model was thus used from the LIME framework, as implemented in the study. As LIME only increases local interpretability, the dropping of data measures are not expected to be of influence on the accuracy concerning the XAI. The input of the LIME explainer is this LightGBM model without missing input values, compared to specific predictions. Due to the method's locality a representative sample of explanations were to be selected. In order to represent the market share division correctly, three different local explanations from the same week were made. One of these explanations represents a high market share, the second one represents an average market share and lastly a case with low market share is explained. The implementation of LIME is done through the lime python package [74].

Decision Tree

To use a decision tree model as an XAI method, the input of this model should not be training data, but rather a different model. The developed LightGBM model, which has been fit to the training data, is used as the decision tree input. The max depth of the decision tree is set to four, keeping the readability of the figure in mind. This choice was made on the basis that the purpose of

the decision tree was to increase model understanding, not to provide an actual prediction. The decision tree's python implementation was done through use of sci-kit learn [91].

Business understanding
In order to understand how business stakeholders view the developed XAI methods structured interviews have been conducted. This method has been chosen, for it can efficiently provide a comparable overview of different stakeholder perspectives [27]. The questions used in the interview consist of both general questions and situational questions, as this combination can help show the interviewee's knowledge and intention [23]. Inspiration for the formulation of the interview questions has been found in earlier XAI audience research [48,26]. In order to keep the interview concise, the questions have been scoped on a specific XAI topic. As there is no definitive research on whether business stakeholders prefer local or global explainability methods, both have been included in the current study. The structured interviews will focus on which of these two classifications of XAI is preferred by the relevant business stakeholders, aiming to provide a first exploration on this topic. The formulated interview questions can be found Appendix A2. Due to the fact that all participants were native Dutch speakers, the interviews have been conducted in Dutch. Therefore, questions and citations used in this paper have been translated to English. As the interviews are of a structured format, the aim during the interview process was to keep all interviews as similar as possible in order to make them comparable.

In order to provide a representative response to this focus area, business stakeholders from different departments have been chosen as interview participants. These departments were: data analysis, supply chain, pricing, store management, buying and customer journey. Their common denominator is that they are all member of a cross-functional team focused on achieving market share increase. To avoid gender biased answers, a mixed selection of males and females have been interviewed. A general overview of the those who participated in the study is provided in the table below.

Participant ID	**Gender**	**Job Department**
A	Male	Supply Chain
B	Female	Store Management
C	Male	Pricing
D	Female	Customer Journey
E	Female	Data Analyst
F	Male	Buying

Table 5: Participant Overview

The interviews have taken place both online and physically, dependent on the participants' schedules. Keeping the goal of similarity in mind, all of the participants were shown the graphs and their textual explanations through the same prepared presentation slides. The participants were then given the time to look at the graphs and read the explanations, to make sure their statements were not made based on mishearing or haste. All interviews were compared afterwards and the participants' answers were considered in the light of their job perspectives. In order to ensure the right statements were used for this comparison, the interviews have been recorded with the consent of the participants. This consent was offered after stating that all recordings would be saved in a company environment and deleted after study completion.

Research Quality
As stated in the research approach, professionals working in the field have been interviewed on their view on the additional value of XAI for their functioning within the business. The participants have been selected from a variety of company departments to avoid creating bias based on job perspective. However, the results from the conducted interviews can not be considered statistically significant nor to be generally applicable. This is due to the low number of participants. The interview results are therefore included as indicative reflections on the discussed XAI methods.

4.3 Results

The explanations from the separate Explainable Artificial Intelligence and the understanding of the business stakeholders will be elaborated upon in the following results section.

Feature Importance
The visualization of traditional feature importance can be done through one simple graph. This graph is shown in the figure below. Within the graph the features and their importance for the model's prediction are shown and ranked based on this importance. The most important features appear to be the amount of visits of the product, the product price of both the company and the most important competitor as well as the difference between, the company's shift in price in comparison to last week, and the delivery time of the company. A possible explanation for the high importance of these features is the fact that they have been experimented with by the company, therefore showing fluctuations that can influence the market share outcome. The three least important features are in what shop the book belongs, its language and whether the book is sold by the company. The former two factors can not be influenced by the company and therefore do not fluctuate over time, which can be a reason for their low importance. Lastly, the assortment of the company is adapted to the market sales. For this reason, not many books will fall within the not for sale category, making it that the feature is the same for almost every product.

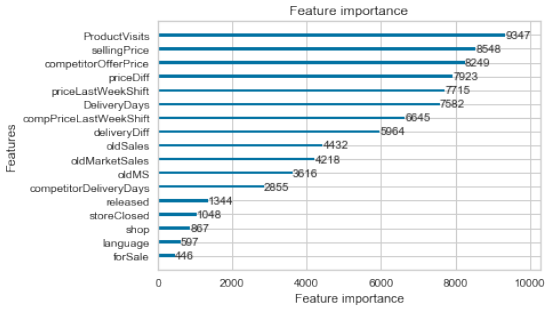


Fig. 18: Feature Importance Global Explanation

Although the graph provides an overview of feature importance, it also has some disadvantages. From this graph alone, no certain statements can be made on how the features are exactly important. This is also reflected in the interview results. Half of the participants (A, B and F) do state that they find this graph easy to read, and therefore easy to understand. However, all of the participants appear to agree that the information that they can gain from the graph is limited. They still do not understand how exactly the model works, nor can they make business decisions upon use of the figure. Participant C captures this sentiment in the following quote:

"This method can only inform me on the degree of impact, but not on whether this impact is positive or negative. I wish it would tell me more."

SHapley Additive exPlanations Framework

The SHAP values globally explaining the LightGBM model predicting market share have been visualized as seen in the figure below. This plot illustrates the importance of features in a few different ways. Firstly, it ranks features from most to least impact on the model's predictions. This impact is also reflected in how much the SHAP value deviates from 0. Secondly, it indicates whether the SHAP value is either negative or positive. The direction of the value represents the direction of the impact that the feature holds. Lastly, the colours in the graph indicate whether the value of the feature itself is high (red) or low (blue) [62].

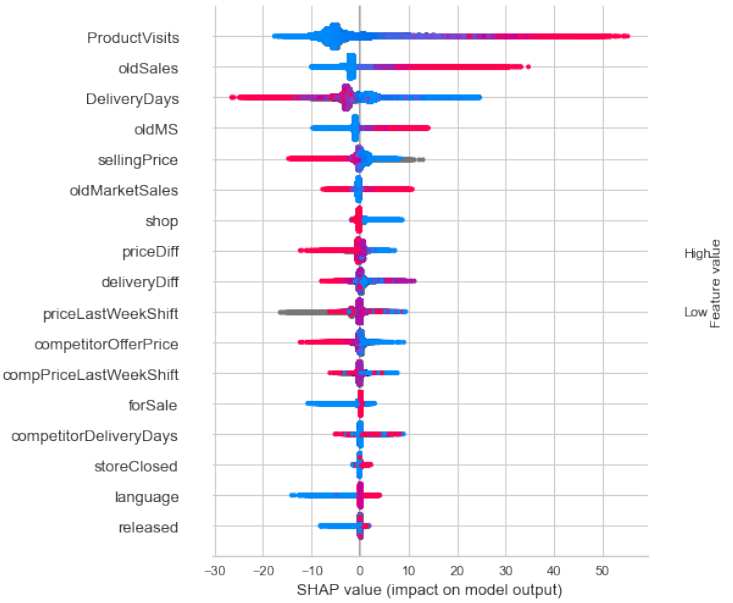


Fig. 19: SHAP Global Explanation

From the SHAP explanation of the LightGBM model some interesting observations follow. The feature with the most impact is product visits. They have a high positive impact on the predicted market share when the visits themselves are high as well. The same logic follows for last week's sales and last week's market share, which are ranked second and fourth in terms of impact. An interesting feature is that of the company's delivery days. The most effect this feature has is when its values are high, having impacted the predicted market share in a negative way. However, low values of market share still hold a considerably large positive effect. Delivery days therefore impact market share prediction in both directions. When comparing this impact to that of selling price, which is ranked as fifth, one can see that price also has this bi-directional impact. However, the negative impact of a high selling price is proportionally bigger than the positive impact of a low selling price. Furthermore, both directions hold less impact than those directions of delivery days do. A last interesting feature, ranked 6th in feature importance, is that of old market sales. Only high previous market sales seem to be indicative of the market share prediction. However, these high values can both impact the market share prediction in a positive and a negative manner.

Features that hold little importance according to the SHAP framework are released, language and storeClosed. The first one meaning whether the book has been released: an unreleased book (value = 0) indicating a low market share. Secondly, the books' language: a non-English book (value $<$ 5) indicating a

low predicted market share. The latter represents whether physical stores were opened: no clear direction.

In the interview results the participants reacted enthusiastically to the insights from Figure 20. All participants described the visualization as both easy to understand and easy to interpret. However, the meaning of the coloured dots was not clear to any of the participants. Participant C describes his understanding of the graph at first glance as:

"I think I understand what the figure means. The further the dots reach, the more impact their related feature has on the output of the model. Dots to the right show a positive effect, while dots to the left point towards a negative effect. The only thing that I can not interpret right away are the colours that are used."

After an explanation, they did understand the figure well. Nonetheless, participants still described the chosen colours are counter-intuitive and noted that a legend next to the graph would have been useful. Furthermore, a few participants noted that while they did understand that the SHAP value reflected impact, it would be nice to know what a SHAP value actually is. This would increase the understanding of the graph, as it is noted on the x-axis.

Not only were the participants asked about their understanding of the graph, they were also asked to evaluate its added value for the business. Aside from participant B, all participants immediately volunteered ways in which they would use the graph to change their decision strategy. Interestingly, most of the interviewees were especially interested in results on features they themselves could control. A prominent benefit of the SHAP visualization appeared to be the direction of the effect, participants valued that they could differentiate between features with a positive and those with a negative impact. Participant A did state that while the graph could be useful, he would only want to act on results that he recognized from the business practice. Nevertheless, overall the reaction to the SHAP figure was that it presents them with an overview of opportunities. Participant E reflects on actions that can be taken with use of the SHAP explanation:

"It is useful to see what features are most important for predicting market share. You can now know which features with a large positive effect you must focus on to keep and on what negatively impacting features you must remove from the business equation."

Interaction Effects

Using the SHAP framework, it is not only possible to look at feature effects, but also at the effects of features interacting. Hence, the interaction values were calculated. An interesting interaction could be between delivery days and selling price, as both are ranked top five in terms of SHAP feature importance and are both features that the company can exercise control over. Their interaction plot

for a singular week is provided in figure 20 below. The values in this dependence plot are coloured based on the height of the selling price value: high values are red and low values are blue. The locations of the data points are determined by the delivery days on the x-axis and their SHAP value on the y-axis.

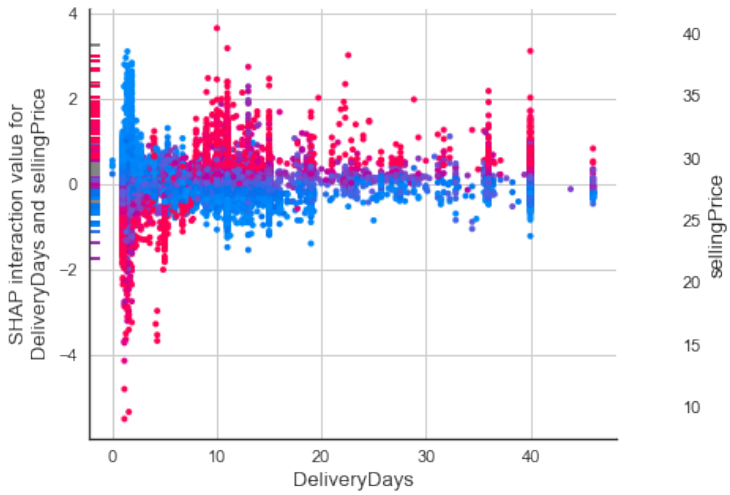


Fig. 20: SHAP Interaction Effect Explanation

Figure 21 shows that the interaction effect of selling price and delivery days in general is not very strong, the majority of the points are spread around the line represeting the SHAP value of zero. This is emphasized by looking at the scale used for the SHAP values in this figure. The axis now ranges from -5 to 5, where in figure 19 the SHAP axis ranges from -30 to 50. This is quite a large difference. The interaction effect indicates that a low price and low delivery days has a small positive impact on market share prediction. However, as stated before these effects are much more diminished than those of the general SHAP explanation. No further conclusions will therefore be connected to this SHAP interaction.

Local Interpretable Model-Agnostic Explanations Framework

In order to showcase the locality of the LIME framework, multiple visualizations have been used, segmented by relevance in the market. These are presented in the figures below. The explanations are shown in three-fold. On the left, the prediction made by the surrogate model is shown, this thus not necessarily mean that it is the exact same prediction as made by the LightGBM model. In the center the feaure importance is illustrated. The importance of the feature can be deducted by looking at (a) the length of the bar, (b) the height of the number next to the bar, or (c) its position: higher features hold more importance.

Furthermore, the centre features indicates whether the impact of a feature is positive (orange) or negative (blue). This is explained by the small legend above. The condition founding the argumentation for this effect is presented above the bar. However, if a feature name is too long, this condition is not entirely readable. Lastly, the right table shows the ranking of ten most important features for the individual prediction and what their values were. Furthermore, the same colours are used to emphasize whether this value had a negative or a positive effect on the predicted market share [26].

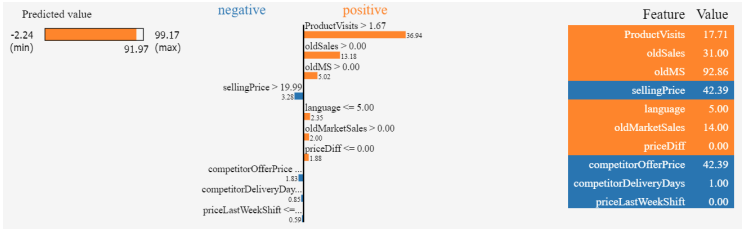


Fig. 21: LIME Top Local Explanation

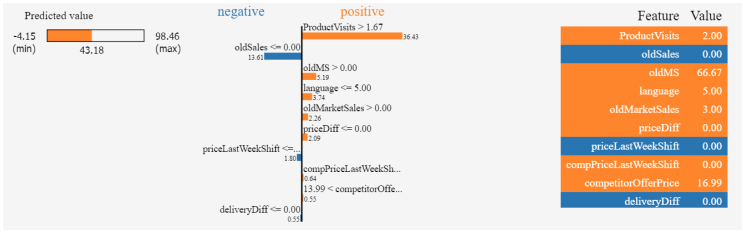


Fig. 22: LIME Middle Local Explanation

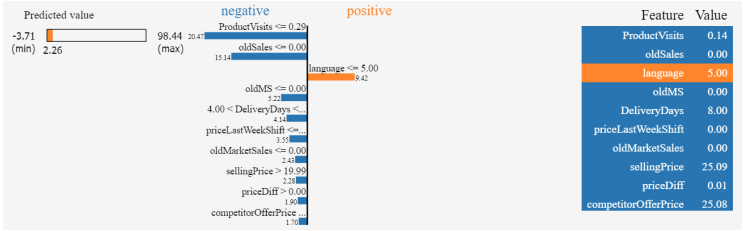


Fig. 23: LIME Tail Local Explanation

Although many interview participants admitted that they did not entirely understand the presented figures, they expressed that the added explanation did clarify all questions that they had. This does exemplify how the LIME framework is not intuitive to understand. Participant C even indicates that he would not want to use the graph due to its counter-intuitiveness. Nonetheless, there were many other enthusiastic responses to the local LIME visualizations. They saw many chances to use segmentation to apply multiple different tactics. Moreover, lessons can be learned this way from books where the company has high market share, and they can be compared to specific books with low market share. This enthusiasm is based primarily on the specificity of these explanations, as is reflected in the statement below.

Participant B: "The ability to zoom in on different predictions provides you with a further step you can take. Now you can specifically connect actions to important assortment and influence their market share."

However, participant A does warn for the segmentation enthusiasm that is found among his colleagues. He says the following:

"Although this clarity on a local level can be very useful, I would be careful to act on this graph. I would not want to make any decisions based on random examples that just fall within a segment."

Little other participants related a similar concern, with only participant F stating that it would be important to use representative local explanations. Although the business stakeholders are enthusiastic about the understanding the LIME provides them with, it remains the question whether they grasp the extent to which these explanations can be generalized.

Alternative Visualization
Aside from the classic LIME visualization, the participants were also presented with an alternative one, shown in figure 24. This figure uses the same LIME data as figure 23, but visualizes this in a different way. In this figure, there is only one graph. Similarly to the previous three figures, the graph ranks the features based on their importance. Their importance can also be seen by the length of the bar. Different than the classic visualization, there is no exact number to represent the importance provided. However, an x-axis representing this importance is included. Besides the importance, the figure shows the condition that determined the direction of their effect. The graph has been fit in such a way that the complete condition is included, which is different than the classic graphs. Lastly, both this and the previous LIME figures use colours to show the direction of the effect. These colours are not the same: while the other LIME visualization used blue and orange, the figure below uses the colours green and red [33].

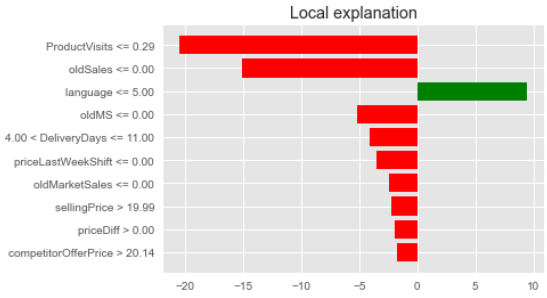


Fig. 24: Alternative LIME

Especially the latter difference made it that during the interviews half of the participants showed a preference towards figure 24, when they were asked to compare both LIME visualizations. They especially experience the colours of this figure as more natural. Participant B comments with:

"In my opinion, this figure provides a better overview. It is nice that it uses red and green, as they are the standard colours for negative and positive. Considering the size of the bar with its x-axis provides me with more insights about the feature importance than the numbers in the other graphs do."

However, not all participants agree with this statement. As this alternative figure is more compact, some worry that with only using this graph other information is lost on them. While participant C, noted that this limited amount of information makes the graph less complex, he is alone in this opinion. Participant A, D and E all state that they think this graph is more difficult to understand, and that they also think this graph is not able to convey as much information as the other LIME figures. Therefore, no clear 'best' visualization of LIME in the business context can be decided upon, due to varying personal favorites.

Decision Tree

The visualization of the developed decision tree is too big to fit in the main text. For this reason, it can be found in Appendix A.3. For the purpose of immediate illustration, a zoomed-in fraction of this output is provided in figure 25. The decision tree is able to track and show the decisions the LightGBM model makes to come to its predictions. Looking at the decision tree, it becomes clear that it shows that the model first considers the amount of visits a product gets. The model is then split into books with two or more visits (21%) and products with less (78.4%). For the minority with multiple visits, the model then considers factors as the last week's company sales, the previous known market share, delivery days, delivery day difference and also comes back to product visits. This indicates that for this group, these factors are decisive for the model in

the early stage. The features that are considered for books with one or no visits are then split again based on their visits. Following this split the model looks at previous sales, delivery days and visits. These features are again important in the final shown layer of the decision tree, where previous market share also plays a part.

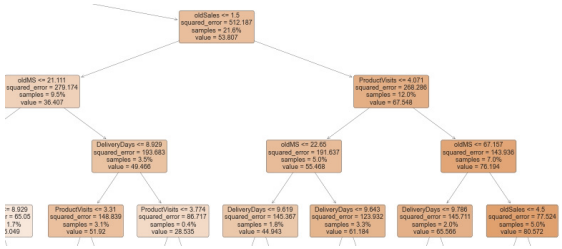


Fig. 25: Fraction of the Decision Tree Explanation

The interview participants had varying responses to the graph visualizing the decision tree. Even though research indicates that a decision tree is straight-forward, and therefore intrinsically interpretable, this does not seem to mean that business stakeholders understand the visualization. Four out of six participants note that they do not understand the graph. They describe the tree as being too complex, too big and that they do not understand how books flow through the tree. Examples of statements related to this sentiment are:

Participant E: “I think the size of this graph makes it that the graph is no longer comprehensible. I would rather have a simple set of rules that the model bases its decisions on.”

Participant B: “I find this figure to be very complex due to the multitude of numbers that are shown.”

However, this does not mean that the stakeholders do not understand the possible worth of a decision tree. Participant D captures the purpose behind the decision tree quite well by stating:

“I find it pleasant to understand what happens inside the model, this way I gain a better understanding of how I should read the model’s output.”

Participants A and C also indicate that they think a similar explanation can add value to the business context. However, both reflect on this worth by stating that it might be used better in different implementations, such as to explain the way the company’s price model makes decisions. With one exception, the participants thus do not consider the decision tree to increase

the interpretability of the LightGBM model. This is primarily due to the complexity of the explanation itself.

Comparison of Explanations
When comparing the results that are produced by the different XAI methods, it becomes clear that they all indicate that product visits have the highest feature importance. Furthermore, the plots that are capable of showing direction show that the effect of visits on market share is positive. The company's last week's sales rank second for almost all methods, with a positive effect. An exception to this is the Feature Importance plot, which considers this feature only nineth most important. This can be explained by the fact that all methods calculate feature importance in a different way. Another interesting discrepancy can be found when comparing the method's ranking of the feature language. While the global methods all indicate that the importance of language can be neglected, the separate local LIME explanations all rank it as a top five feature. The importance of delivery days is another variable that not all methods agree on. While the decision tree and the SHAP framework both rank it third, this is not as much the case for the feature importance plot and two of the LIME explanations. The latter seem to adhere more value to price-related features. However, the tail LIME explanation does show the negative impact of a high delivery days value, which is comparable to the SHAP framework.

Business Understanding
Besides their opinion and understanding of the separate used XAI methods, the interview participants were presented with some comparative questions. These questions aimed to understand the preferences of the participants, as well as to gain knowledge on how they would use their preferred figure. Before any of the participants were showed any figure, they were explained the difference between local and global XAI. Based on this explanation they were asked which one of these explanation types they think has the most added value within a business context. All participants responded in a similar way, they showed a preference towards a global explanation. This preference seems to be related to the volume of the books assortment, as participant F expresses in his answer:

"Global explanations sound the most useful, because we have so many unique books: around 10 million. Therefore it is most important what happens in general, because you can assume most products will adhere to this normal pattern. If we were to know what this pattern entails, then we could make an estimation which products would follow an alternate pattern."

All other participants provide a similar answer, using the same reasoning. Only participant D says that she can imagine that a local explanation would make a lot of sense, as one can trace this explanation back to a specific product. Nonetheless, she still expresses a preference towards global. However,

after the participants had seen all the different explanations, they were asked which one of those would have the most added worth within the business context. This time, not all participants immediately opted for the graphs showing global explanations. Participants A, C and E all keep their originally expressed preference, by stating that they think the most added value can be found with the SHAP figure. As Participant A argues:

"The most value can be added by Figure 19. It simultaneously provides the most insight while also not requiring you too look at too many details. When looking at these graphs I primarily want to know which features, broadly speaking, have the most impact. This I think is most represented in that figure"

Nonetheless, this way of thinking is not shared by participants B, D and F. After comparing all graphs, they state that the global explanations might be a little too broad. Furthermore, participant B states that her preference is the alternative LIME visualization, as she believes that this graph will be understood by any business stakeholder. Whether this is true is unsure, as the other two participants stated that the classic LIME visualization holds their preference. Participant D explains her choice for the local LIME explanations with the following statement:

"I think that in the end this local explanation [ed: the classic LIME visualization] will show to be the most useful. If the explanations can be used to represent a segment of products this might easier lead to concrete points of action. I am unsure how big the impact of these actions will be though."

Considering these two contrasting preferences, it remains unclear what XAI explanation is preferred by business stakeholders. However, a possible explanation can be found when considering the participants' jobs. The participants favouring the global explanation are those that work in departments that work in a broad manner, not focusing on specific products, but implementing changes that will effect the product group overall. The opposite is true for the group that states that they prefer the local explanations. The buying, store management and customer journey departments in their operational work often have to focus on a specific book. When making changes, they often make small ones to make specific high-performing products stand out. A result that thus seems to follow from the comparative questions is that business stakeholders prefer an XAI explanation that is related the most closely to their way of working.

4.4 Explainable Artificial Intelligence Conclusion

Interviews conducted with business stakeholders, showed that the graphs provided them with more understanding than they had before the XAI methods. For all figures, besides the one showing only feature importance, they

did need explanations to correctly interpret the graph. Especially the colour usage in the visualizations was unclear to them, describing it as "counter-intuitive". Out of the more complex visualizations, the SHAP figure received the least amount of comments on its complexity. This is similar to the existing literature cited in the theoretical framework, which describes the SHAP framework as very intuitive. The LIME understanding of the LIME visualizations differed by participant, some thought it was very intuitive, while others complained about its complexity. The interview results showed that the decision tree is considered by the stakeholders to be the most difficult to interpret. Here the interview results differ from general literature on decision trees, which are characterized by their 'straight-forward nature'. The non-technical background of the stakeholders might play a part in this disparity.

In terms of relevance for the business context, the participants were split in their opinion. In general, they were enthusiastic about the graphs presented to them. Especially the LIME and SHAP explanations led to an increased sense of understanding the model's predictions and the features. However, no conclusion on the participants' method preference can be drawn. One side of the participants preferred the global nature of the SHAP explanation, arguing that its output could help them understand best what is happening overall and points them in the direction of which feature change can make the most impact. The other side suggested that with use of the LIME explanation it is possible to choose representative samples to segment the company's assortment. This could help the company to employ more targeted actions, however the uncertainty on the size of the impact of these actions remains. However, the participants might not take the manual labor that is needed to find these representative samples into account. This extra work is a disadvantage of a local framework, as the literature on the subject stated.

5 Conclusions and Discussion

Explainable Artificial Intelligence is able to contribute to the prediction of market share in the field of e-commerce by explaining to business stakeholders what the importance of features is for the model making this prediction. These insights can help the stakeholders to evaluate their current decisions, and to adjust accordingly. This conclusion of the study was reached by considering two hypotheses in response to the research question: How can XAI contribute to the prediction of market share in the field of e-commerce?

The first hypothesis covered the prediction aspect of the research question, stating that a machine learning model was capable of predicting the market share of international book sales done by online parties. To evaluate this hypothesis several machine learning models were experimented with, and tested in terms of the accuracy of their predictions. These models were selected based on whether they were suitable for a supervised regression problem. Out of these models, the Light Gradient Boosting Machine (LightGBM) model performed the best in

predicting market share. This was reflected in the amount of data the LightGBM model could process and the height of accuracy of the predictions, which sufficed for the goal of this study. Furthermore, the results showed that if more data were to collected, especially on competitors, this accuracy would increase even further, highlighting the predictive abilities of the model. This hypothesis is accepted based on the study results.

Nevertheless, there are some limitations to the acceptance of this hypothesis. Firstly, not all events happening in the market can be captured by the model, due to the limited amount of data available. This is reflected in two ways. Firstly, the data collected does not include possible promotions done by the competitor. Moreover, other aspects of the competitor data have also been been found to be missing in the dataset, while these aspects are available for the company data. Together, these aspects of missing data lead to a decrease in accuracy, as the LightGBM model does not have all possible information that can help make a prediction. Another point of discussion is that of seasonality. No seasonality effects were taken into account for the current study. However, this is largely due to practical limitations as well as a small scope of included data. Therefore, possible seasonality effects could be missing from the model.

The second hypothesis attends to the Explainable Artificial Intelligence (XAI) aspect of the research question, stating that the use of XAI can help a business to determine features that are important for a predictive machine learning model. This was tested in twofold. Firstly, several XAI methods and their explanations of the LightGBM model were explored. These methods were selected based on whether they could provide post-hoc interpretability in a model-agnostic way. However, both local and global explanations were included as well as methods that provided either a feature summary or surrogate intrinsic interpretability. All of the methods were capable of visualizing features that were important for market share prediction. However, they all illustrate this importance in a different way. These difference are covered by the second test of the hypothesis. The relevance of the methods were compared through interviews of business stakeholders working at the study's host company. The interview results demonstrated that the participants were enthusiastic about the use of XAI, as it provided them with more in-depth insights into the predictions. Especially the methods that were able to show direction of effect were received with positive responses. However, the participants were not unanimous in their preference of a specific XAI method. Business stakeholders appeared to favour an explanation that was similar to their way of working, those who primarily work in a way that effects the product overall prefer a global XAI method, while those that work in a product-based way state a preference toward local explanations. The second hypothesis is accepted based on the study results.

However, when accepting this hypothesis, one should note the restraints of the hypothesis within this study. As noted when discussing the previous hypothesis, the model analyzed by the XAI methods does not capture the entirety of the market. Consequently, neither does the XAI method. It is important to note

that the method explains the model, not the market behavior, and these two should not be confused. Nonetheless, as the LightGBM model does predict the market sufficiently, it can be stated that the XAI explanations do approximate the market. Another restraint for accepting the hypothesis is that the interview results are based off of a small sample of the desired population. This makes that the interview results, used to accept the hypothesis, can not be generalized for the business stakeholders as a population.

5.1 Future Research

Based on the successes and shortcomings of the conducted study a few recommendations are made for future research. These recommendations are to apply more focus to seasonality effects, to expand the number of interview participants and to explore ways in which the results can be implemented within businesses in a scalable way.

Seasonality
As was noted in the paragraph on seasonality, no seasonality was found in the data used for this study. However, this could be due to the limited span of time covered by this data. As only one year of data was included in the study, it was not possible to find any yearly seasonality. Furthermore, the Facebook-Prophet model, developed to handle time series data and different types of seasonality, could not be used to make the market share predictions. As stated before, due to practical limitations this model was not experimented with in the study. For this reason, the first recommendation for future research is to consider possible seasonality effects on market share more closely. In order to do so, additional data covering multiple years should be selected. Consequently, yearly seasonality can be evaluated. Another way to explore seasonality is by predicting market share through the Facebook-Prophet model.

Interviews
From the interviews some interesting results on perspectives of business stakeholders on different XAI explanations followed. However, as touched upon briefly, these results can not be generalized to a larger business stakeholder population. This is due to the small sample of participants that were interviewed. This can make implementation of the results difficult. A recommendation for further research is to expand this sample, in order ensure that it is more representative of researched population. Furthermore, this expansion can lead to more qualitative insights on the thoughts of different stakeholders, as well as further insights into whether the effect of way of working on XAI preference is reproduced.

Result Implementation
The current study has shown that the use of XAI in e-commerce can lead to business stakeholders gaining more knowledge on the way machine learning can contribute to making business decisions. Furthermore, XAI can provide insights

on the dynamics within a domain such as market share, providing the company with a direction for their actions. Nonetheless, there is no clear way to implement these positive results. For the study, stakeholders were shown the XAI figures on powerpoint slides. This is not a scalable way of working, especially not when multiple departments and product groups would get involved. Additionally, in order to have the found results operate in the long run, the implementation would be required to integrate the other aspects that are recommend for further research. The first requirement is that the implementation would be able to handle the feedback from a large group of stakeholders, collecting data on the preferences of different departments. Furthermore, the implementation should be synchronized with the collection of additional data to continuously be updating. Ideally, more data would be added to the model, extending the coverage of the market. Connected to the extension of data, another implementation requirement is that of storage. This storage should function efficiently, as a lot of data should be stored. This is both due to the large amount of products the book market holds, as well as the indicated need for a large time span of collected data to measure seasonality effects. A recommendation for further research is thus to experiment with different ways to implement XAI methods, aiming to find a way that functions within a big organization.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (xai). IEEE access **6**, 52138–52160 (2018)
2. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
3. Al-Hameedi, A.T.T., Alkinani, H.H., Dunn-Norman, S., Flori, R.E., Hilgedick, S.A., Amer, A.S., Alsaba, M.: Mud loss estimation using machine learning approach. Journal of Petroleum Exploration and Production Technology **9**(2), 1339–1354 (2019)
4. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., Aljaaf, A.J.: A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science pp. 3–21 (2020)
5. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al.: Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information Fusion **58**, 82–115 (2020)
6. Ayodele, T.O.: Types of machine learning algorithms. New advances in machine learning **3**, 19–48 (2010)
7. Bass, F.M., Pilon, T.L.: A stochastic brand choice framework for econometric modeling of time series market share behavior. Journal of Marketing Research **17**(4), 486–497 (1980)
8. Bisong, E.: Ensemble methods. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform, pp. 269–286. Springer (2019)
9. Bisong, E.: Introduction to scikit-learn. In: Building machine learning and deep learning models on Google cloud platform, pp. 215–229. Springer (2019)
10. Blundell, R., Griffith, R., Van Reenen, J.: Market share, market value and innovation in a panel of british manufacturing firms. The review of economic studies **66**(3), 529–554 (1999)
11. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
12. Broniatowski, D.A., et al.: Psychological foundations of explainability and interpretability in artificial intelligence. NIST, Tech. Rep (2021)
13. Brooks, G.R.: Defining market boundaries. Strategic Management Journal **16**(7), 535–549 (1995)
14. Buzzell, R.D., Gale, B.T., Sultan, R.G.: Market share-a key to profitability. Harvard business review **53**(1), 97–106 (1975)
15. Cao, N.: Explainable artificial intelligence for customer churning prediction in banking
16. Carlton, D.W.: Market definition: Use and abuse. Competition Policy International **3**(1) (2007)
17. Carvalho, D.V., Pereira, E.M., Cardoso, J.S.: Machine learning interpretability: A survey on methods and metrics. Electronics **8**(8), 832 (2019)

18. Casalicchio, G., Molnar, C., Bischl, B.: Visualizing the feature importance for black box models. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 655–670. Springer (2018)
19. Chai, T., Draxler, R.R.: Root mean square error (rmse) or mean absolute error (mae). Geoscientific Model Development Discussions **7**(1), 1525–1534 (2014)
20. Chaturvedi, A., Dhariwal, A., Patel, M.: Study on prediction of airfares based on xgboost and light gbm machine learning algorithms. International Research Journal of Modernization in Engineering Technology and Science **2**(4) (2020)
21. Chen, I.F., Lu, C.J.: Sales forecasting by combining clustering and machine-learning techniques for computer retailing. Neural Computing and Applications **28**(9), 2633–2647 (2017)
22. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., et al.: Xgboost: extreme gradient boosting. R package version 0.4-2 **1**(4), 1–4 (2015)
23. Conway, J.M., Peneno, G.M.: Comparing structured interview question types: Construct validity and applicant reactions. Journal of Business and Psychology **13**(4), 485–506 (1999)
24. Cooper, L.G.: Market-share models. Handbooks in operations research and management science **5**, 259–314 (1993)
25. Das, A., Rad, P.: Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv preprint arXiv:2006.11371 (2020)
26. Dieber, J., Kirrane, S.: Why model why? assessing the strengths and limitations of lime. arXiv preprint arXiv:2012.00093 (2020)
27. Doloi, H.K.: Understanding stakeholders' perspective of cost estimation in project management. International journal of project management **29**(5), 622–636 (2011)
28. Du, M., Liu, N., Hu, X.: Techniques for interpretable machine learning. Communications of the ACM **63**(1), 68–77 (2019)
29. von Eschenbach, W.J.: Transparency and the black box problem: Why we do not trust ai. Philosophy & Technology **34**(4), 1607–1622 (2021)
30. Eveleens, J.L., Verhoef, C.: Quantifying it forecast quality. Science of computer programming **74**(11-12), 934–988 (2009)
31. Ferrier, W.J., Smith, K.G., Grimm, C.M.: The role of competitive action in market share erosion and industry dethronement: A study of industry leaders and challengers. Academy of management journal **42**(4), 372–388 (1999)
32. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics pp. 1189–1232 (2001)
33. Garreau, D., von Luxburg, U.: Looking deeper into tabular lime. arXiv preprint arXiv:2008.11092 (2020)
34. Glock, A.C.: Explaining a random forest with the difference of two arima models in an industrial fault detection scenario. Procedia Computer Science **180**, 476–481 (2021)
35. Grömping, U.: Variable importance assessment in regression: linear regression versus random forest. The American Statistician **63**(4), 308–319 (2009)
36. GrowthFromKnowledge: (2022), https://gfkpanel.nl/home
37. Gulowaty, B., Woźniak, M.: Extracting interpretable decision tree ensemble from random forest. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2021)

38. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z.: Xai—explainable artificial intelligence. Science robotics **4**(37), eaay7120 (2019)
39. Gupta, B., Rawat, A., Jain, A., Arora, A., Dhami, N.: Analysis of various decision tree algorithms for classification in data mining. International Journal of Computer Applications **163**(8), 15–19 (2017)
40. Hakkoum, H., Idri, A., Abnane, I.: Artificial neural networks interpretation using lime for breast cancer diagnosis. In: World Conference on Information Systems and Technologies. pp. 15–24. Springer (2020)
41. Hancock, J., Khoshgoftaar, T.M.: Leveraging lightgbm for categorical big data. In: 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService). pp. 149–154. IEEE (2021)
42. Hapfelmeier, A., Ulm, K.: Variable selection by random forests using data with missing values. Computational Statistics & Data Analysis **80**, 129–139 (2014)
43. Heij, C., Heij, C., de Boer, P., Franses, P.H., Kloek, T., van Dijk, H.K., et al.: Econometric methods with applications in business and economics. Oxford University Press (2004)
44. Ho, S.L., Xie, M.: The use of arima models for reliability forecasting and analysis. Computers & industrial engineering **35**(1-2), 213–216 (1998)
45. Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., Baesens, B.: An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decision Support Systems **51**(1), 141–154 (2011)
46. Islam, S.R., Eberle, W., Ghafoor, S.K., Ahmed, M.: Explainable artificial intelligence approaches: A survey. arXiv preprint arXiv:2101.09429 (2021)
47. Jiang, H., Senge, E.: On two xai cultures: A case study of non-technical explanations in deployed ai system. arXiv preprint arXiv:2112.01016 (2021)
48. Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan, J.: Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–14 (2020)
49. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems **30** (2017)
50. Keane, M.T., Kenny, E.M.: How case-based reasoning explains neural networks: A theoretical analysis of xai using post-hoc explanation-by-example from a survey of ann-cbr twin-systems. In: International Conference on Case-Based Reasoning. pp. 155–171. Springer (2019)
51. Khaidem, L., Saha, S., Dey, S.R.: Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003 (2016)
52. Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., Baum, K.: What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. Artificial Intelligence **296**, 103473 (2021)
53. Li, K., Xu, H., Liu, X.: Analysis and visualization of accidents severity based on lightgbm-tpe. Chaos, Solitons & Fractals **157**, 111987 (2022)
54. Li, L.: Rolling Window Time Series Prediction using MapReduce. Ph.D. thesis, University of Sydney (2014)

55. Li, L., Qiao, J., Yu, G., Wang, L., Li, H.Y., Liao, C., Zhu, Z.: Interpretable tree-based ensemble model for predicting beach water quality. Water Research **211**, 118078 (2022)
56. Liao, Q.V., Varshney, K.R.: Human-centered explainable ai (xai): From algorithms to user experiences. arXiv preprint arXiv:2110.10790 (2021)
57. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: From local explanations to global understanding with explainable ai for trees. Nature machine intelligence **2**(1), 56–67 (2020)
58. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)
59. Mahbooba, B., Timilsina, M., Sahal, R., Serrano, M.: Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model. Complexity **2021** (2021)
60. Manasa, J., Gupta, R., Narahari, N.: Machine learning based predicting house prices using regression techniques. In: 2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA). pp. 624–630. IEEE (2020)
61. Marcílio, W.E., Eler, D.M.: From explanations to feature selection: assessing shap values as feature selection mechanism. In: 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI). pp. 340–347. Ieee (2020)
62. Meng, Y., Yang, N., Qian, Z., Zhang, G.: What makes an online review more helpful: an interpretation framework using xgboost and shap values. Journal of Theoretical and Applied Electronic Commerce Research **16**(3), 466–490 (2020)
63. Microsoft: Microsoft/lightgbm: A fast, distributed, high performance gradient boosting (gbt, gbdt, gbrt, gbm or mart) framework based on decision tree algorithms, used for ranking, classification and many other machine learning tasks. (Apr 2017), https://github.com/microsoft/LightGBM
64. Microsoft: Lightgbm 3.3.2.99 documentation (2022), https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.plot_importance.html
65. Minastireanu, E.A., Mesnita, G.: Light gbm machine learning algorithm to online click fraud detection. J. Inform. Assur. Cybersecur **2019** (2019)
66. Molnar, C.: Interpretable machine learning (2020)
67. Nasteski, V.: An overview of the supervised machine learning methods. Horizons. b **4**, 51–62 (2017)
68. Nielsen: Nielseniq and gorillas sign a cooperative relationship to enhance "quick commerce" read and insights (Jul 2022), https://nielseniq.com/global/en/news-center/2022/nielseniq-and-gorillas-sign-a-cooperative-/relationship-to-enhance-quick-commerce-read-and-insights/
69. Notenboom, A., Schrijvershof, C., Goudriaan, R.: Evaluatie van de wet op de vaste boekenprijs. Een kwantitatieve analyse. Ape rapport (650) (2009)
70. Nwogu, E.C., Iwueze, I.S., Nlebedim, V.U.: Some tests for seasonality in time series data. Journal of Modern Applied Statistical Methods **15**(2), 24 (2016)
71. Pandey, A., Rastogi, V., Singh, S.: Car's selling price prediction using random forest machine learning algorithm. In: 5th International Conference on Next Generation Computing Technologies (NGCT-2019) (2020)

72. Park, J.H., Jo, H.S., Lee, S.H., Oh, S.W., Na, M.G.: A reliable intelligent diagnostic assistant for nuclear power plants using explainable artificial intelligence of gru-ae, lightgbm and shap. Nuclear Engineering and Technology **54**(4), 1271–1287 (2022)
73. Pavlyshenko, B.M.: Machine-learning models for sales time series forecasting. Data **4**(1), 15 (2019)
74. Poduska, J.: Shap and lime python libraries: Part 1–great explainers, with pros and cons to both (2018)
75. Pradeepta, M.: Practical explainable ai using python (2022)
76. Rai, A.: Explainable ai: From black box to glass box. Journal of the Academy of Marketing Science **48**(1), 137–141 (2020)
77. Ramon, Y., Vermeire, T., Toubia, O., Martens, D., Evgeniou, T.: Understanding consumer preferences for explanations generated by xai algorithms. arXiv preprint arXiv:2107.02624 (2021)
78. Reitermanova, Z., et al.: Data splitting. In: WDS. vol. 10, pp. 31–36. Matfyzpress Prague (2010)
79. Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1135–1144 (2016)
80. Schnaubelt, M.: A comparison of machine learning model validation schemes for non-stationary time series data. Tech. rep., FAU Discussion Papers in Economics (2019)
81. Schober, P., Boer, C., Schwarte, L.A.: Correlation coefficients: appropriate use and interpretation. Anesthesia & Analgesia **126**(5), 1763–1768 (2018)
82. Seabold, S., Perktold, J.: Statsmodels: Econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference. vol. 57, pp. 10–25080. Austin, TX (2010)
83. Shumway, R.H., Stoffer, D.S., Stoffer, D.S.: Time series analysis and its applications, vol. 3. Springer (2000)
84. Skrivanek, S.: The use of dummy variables in regression analysis. More Steam, LLC (2009)
85. Stadnytska, T.: Deterministic or stochastic trend: Decision on the basis of the augmented dickey-fuller test. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences **6**(2), 83 (2010)
86. Stank, T.P., Goldsby, T.J., Vickery, S.K., Savitskie, K.: Logistics service performance: estimating its influence on market share. Journal of business logistics **24**(1), 27–55 (2003)
87. Stepanek, H.: Thinking in Pandas. Springer (2020)
88. Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., Cilar, L.: Interpretability of machine learning-based prediction models in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**(5), e1379 (2020)
89. Sun, X., Liu, M., Sima, Z.: A novel cryptocurrency price trend forecasting model based on lightgbm. Finance Research Letters **32**, 101084 (2020)

90. Swanson, N.R., White, H.: Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. International journal of Forecasting **13**(4), 439–461 (1997)
91. Tran, M.K., Panchal, S., Chauhan, V., Brahmbhatt, N., Mevawalla, A., Fraser, R., Fowler, M.: Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery. International Journal of Energy Research **46**(2), 786–794 (2022)
92. Tso, G.K., Yau, K.K.: Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy **32**(9), 1761–1768 (2007)
93. Uyanık, G.K., Güler, N.: A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences **106**, 234–240 (2013)
94. Vermeire, T., Laugel, T., Renard, X., Martens, D., Detyniecki, M.: How to choose an explainability method? towards a methodical implementation of xai in practice. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 521–533. Springer (2021)
95. Wang, S.C.: Artificial neural network. In: Interdisciplinary computing in java programming, pp. 81–100. Springer (2003)
96. Weng, T., Liu, W., Xiao, J.: Supply chain sales forecasting based on lightgbm and lstm combination model. Industrial Management & Data Systems **120**(2), 265–279 (2019)
97. Xu, X.: Price dynamics in corn cash and futures markets: cointegration, causality, and forecasting through a rolling window approach. Financial Markets and Portfolio Management **33**(2), 155–181 (2019)
98. Xu, Y., Kong, Q.J., Klette, R., Liu, Y.: Accurate and interpretable bayesian mars for traffic flow prediction. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS **15**(6), 2457 (2014)
99. Yang, L., Shami, A.: On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing **415**, 295–316 (2020)
100. Zednik, C.: Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy & technology **34**(2), 265–288 (2021)
101. Zhang, G.P.: Time series forecasting using a hybrid arima and neural network model. Neurocomputing **50**, 159–175 (2003)
102. Zunic, E., Korjenic, K., Hodzic, K., Donko, D.: Application of facebook's prophet algorithm for successful sales forecasting based on real-world data. arXiv preprint arXiv:2005.07575 (2020)

A Appendix

A.1 Extended Exploratory Data Analysis

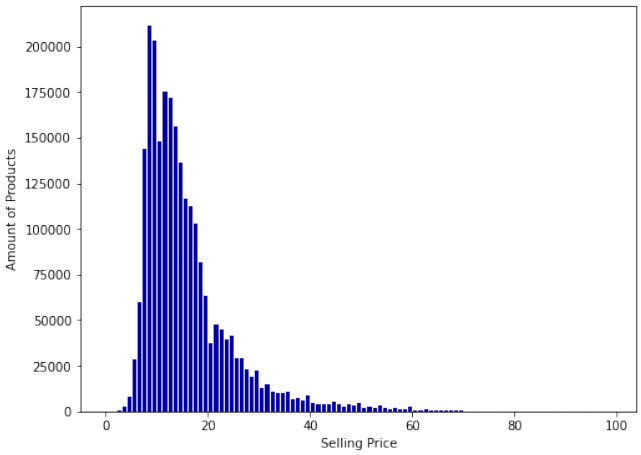


Fig. 26: Distribution of Company Selling Price

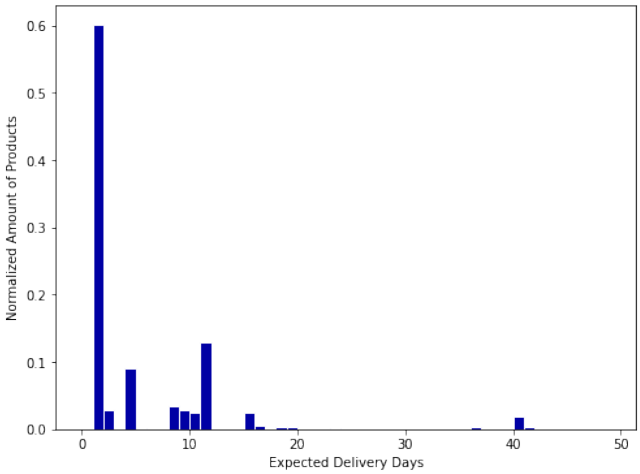


Fig. 27: Distribution of Company Delivery Days

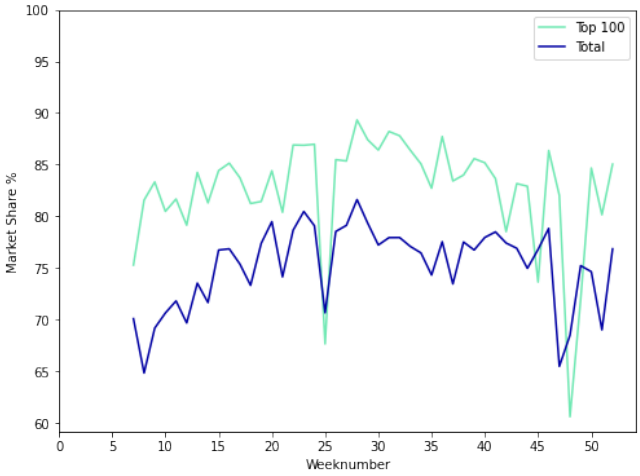


Fig. 28: Top 100 vs. Total Comparison

A.2 Interview Questions

Please note that the interviews were held in the native language of the participants: Dutch. The questions provided below are translations of the questions used in these interviews.

1. What is your function within the company?

Explainable Artificial Intelligence methods are able to provide global as well as local explanations. Global in this case means that the explanation is focused on how the model makes general decisions and which factors are important in this decision-making. Local explanations are those that for a individual prediction (thus focused on a specific GID in a specific week) explain why the model made this specific prediction.

2. Which of these two methods appears to be the most appropriate in a business context? Why?

3a. [Shows Figure 19] This is an example of a global explanation. Do you understand the meaning of this graph?

In the graph factors are ranked based on their importance for the model's prediction. Visits appear to be very important, while whether a book is released not so much. Furthermore, you see a 0-axis line in the graph. Points left of this line represent a negative impact on market share and point right of the line represent a positive impact. Furthermore, there are two colors in the graph. Red represents a high value of that factor and blue represents a low value.

3b. Would this global explanation support you in your decision-making? Why?

4 [Shows Figure 18] This is a graph generated by the model itself. How would you compare this and the previous graph in terms of convenience.

5a. [Shows Figures 21, 22 and 23] These are examples of local explanations, divided over top - middle - tail. Do you understand the meaning of this graph?

In this graph you can find the predicted market share value in the upper left. In the middle you find a ranking of the importance of the factors for this specific model prediction. The weight of this factor can be recognized by looking at the values beneath the bar of the length of the bar. The bars are either orange (positively impacting market share) or blue (negatively impacting market share. Furthermore, all factors have been given a condition that helped the model make a decision en are true in the current prediction. On the right you can find the most important factors for this individual prediction, what the values of those factors were for this specific book and whether these had a positive or negative impact.

5b. Would this local explanation support you in your decision-making? Why?

5c. [Shows Figure 24] This is the same graph but in a different format. How would you compare this graph to the blue/orange one from Figure 23.

6. [Shows Figure 29] These are the first few layers of a decision tree, illustrating how the model made choices. Do you think this model is relevant in a business context?

7a. Which of the graphs from this interview did you find to add the most value within a business?

7b. Case: Imagine that you want to figure out how you can improve the company's market share within the online book market. Which graph would you select to answer this question and what from the graph would you use?

A.3 Decision Tree Output

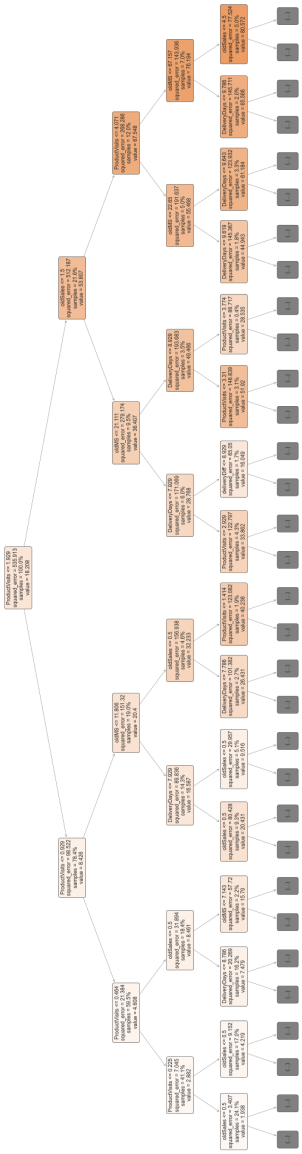


Fig. 29: Complete Decision Tree Explanation