VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

SOCIAL MEDIA PERFORMANCE AND EVALUATION: AN APPROACH TO BUSINESS ANALYTICS CONCEPT

Genesis Sembiring Depari, Fakultas Ekonomi Universitas Pelita Harapan Medan
email : genesis.sembiring@uph.edu
Adat Muli Peranginangin, STIE Surya Nusantara
email : adatperanginangin09@gmail.com
Albert Owen, Fakultas Ekonomi Universitas Pelita Harapan Medan
Felix Emmanuel, Tanny Fakultas Ekonomi Universitas Pelita Harapan Medan
Luis Jonatan, Fakultas Ekonomi Universitas Pelita Harapan Medan
Winnie Lauren, Fakultas Ekonomi Universitas Pelita Harapan Medan

ABSTRACT

The growth of the internet has resulted in the digitalization of data, which has led to the emergence of big data opportunities. Significant amounts of digital data leave traces of what customers see, read, do, and judge, as well as information about their interests and preferences, resulting in a large amount of data that may be mined for learning experiences. Data mining, statistical algorithms, and machine learning approaches are used in descriptive, predictive, and prescriptive analytics to analyze, forecast, and optimize what is the most take effect, future trends, events, and behaviors based on various data types. A decision support system is widely demanded in tackling this problem, especially in understanding the interactions based on the type, and time from the Facebook post about branding data sets. This work attempts to offer descriptive, predictive, and prescriptive analytics to determine whether a post is worth paying for and promoting. This study is sought for deeper observations of posts on Facebook that get a lot of interaction and loyal users by the best algorithm compared with naive Bayes and decision tree which is using Random Forest with 90.35 % accuracy.

Keywords : Business Analytics, Random Forest, Support Vector Machine, Naive Bayes, Decision Tree, Data Mining.

I. INTRODUCTION

In the current era, the development of social media in all over the world is very rapid, various information can be received through social media. In this technology era, a new trend of using Social Media or Social Networking Sites (SNSs) has been recently highlighted (Kamnoetsin, 2015). The current use of social media is very concerning because many social media users prioritize social media over other needs. It's as if electronic goods and quotas are primary needs at this time.

Internet access and usage in the world has been proliferating year by year, with approximately 1.11 billion users in 2007, 1.67 billion in 2009, and 1.97 billion in 2010 (Xanthidis, 2008) Social media is a medium for socializing with each other and done online which allows humans to interact with each other without being limited by space and time. One of the social media that is widely used in Thailand is Facebook.

Facebook is a platform that own by Meta first, which is an American social media and online social networking website (Kraus ,2020). Mark Zuckerberg and a group of Harvard College students founded Facebook in 2004. Because of the widespread usage of social media, social media marketing has become one of the most important aspects of global marketing. The majority of 94 percent of enterprises in the globe use social media for marketing objectives, and demonstrating this (Basri, 2017).

Companies quickly recognized the power of using Internet-based social networks to influence clients, and they began implementing social media marketing communication into their business plans. Measuring the impact of advertising is a crucial component of a global social media strategy (Lariscy et al., 2009). Several research have looked into the relationship between online articles on social networks and their impact as measured by user interactions. When opting to interact via social media and adjusting the marketing of products and services, a system that can forecast the impact of specific published posts can be a huge help. Advertising managers may make informed decisions on the receptiveness of the posts they produced, aligning efforts to maximize the impact of posts while taking advantage of the predictions and using prescriptive analysis. Furthermore, it has been established that social media publications are strongly linked to brand building (Edosomwan et al., 2011). As a result, the prediction tool described in this research could help managers make better judgments about marketing awareness.

Data mining is a fascinating method for obtaining predictive knowledge from unstructured data (Turban, 2011). Its usage in social media has been investigated, particularly for evaluating market trends based on user input (Trainor, 2014). The majority of the research, on the other hand, concentrated on a reactive examination of what people are saying on social media, with an emphasis on acquiring information from various network groups or even personal posts (e.g., Bianchi and Andrews, 2015). We concentrated on forecasting the impact of individual postings being published on a social media network company's page. Several variables linked to consumer visualizations and interactions are used to assess the impact. The predicted knowledge discovered can help managers decide whether or not to publish each post.

II. LITERATURE REVIEW

Business Analytics

Business analytics is a revolution that cannot be overlooked. Business analytics is all about extracting value from data. Rather than being referred to as the "sludge of the digital era," data is now referred to be "the new oil." While data may be used for a variety of objectives, including spotting new opportunities, identifying market niches, and generating new products and services, it is infamously amorphous and difficult to value extract (Acito, 2014).

Self - service business analytics tools, artificial intelligence (AI), and cloud data management software have enabled basically everybody (regardless of skill level) to analyze and visualize patterns in real time, spot company problems, and make informed business decisions in the recent decade. Computer scientists are no longer the only ones who can perform business analysis. Finance, manufacturing, healthcare, sales, marketing, supply chain, and operations are just a few of the industries that use it. Business analytics is required if you want to gain insights from your data.

Business analytics skills are now taught in several degree programs and are a significant aspect of many of them. Students in business learn to apply their knowledge in real - world scenarios such as corporate operations, where they identify essential Key performance metrics and use a data - driven approach.

Three Domain of Analytics

Descriptive Analytics

Descriptive analysis is one of the most complex, adaptable, and commonly utilized methods In the field of sensory analysis. Many strategies have been created throughout the years to satisfy various purposes and uses, each with its own set of benefits and drawbacks. Flavour profiling, texture profiling, the SpectrumTM Method, and quantitative descriptive analysis are examples of traditional methodologies. Because of the intensive training necessary, these take longer and cost more money, but the data supplied is more thorough and detailed. Free Choice Profiling, Flash Profiling, sorting, projective mapping, and Polarised Sensory Positioning are some of the more recent approaches.

These are easier to use because they don't require any training, but the data processing is more complicated and the findings are less thorough. There are other completely qualitative methods, such as Tick – All – That - Apply and open questioning, to quantify temporal changes in products, such as

Jurnal AC USI AccUSI Journal Of Accounting USI
VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

continuous Time-Intensity and Temporal Dominance of Sensation. Many businesses utilize their own, custom-made descriptive ways (Kemp, 2018).

- Mean
  The average of a group of scores is called the mean. Add up the points and divide by the total number of points. When population samples are tiny, the mean is susceptible to extreme scores.
  Mean (m) = sum of the terms / numbers of terms

- Median
  The median is the point where half of the scores are higher and half are lower. Medians are less sensitive to high scores and, especially for smaller sample numbers, are probably a better indicator of where the middle of the class is performing.

$$\text{N is odd; Median} = \left(\frac{n+1}{2}\right)^n$$

$$\text{Or, when N is Even; Median} = \frac{\left(\frac{n}{2}\right)^n + \left(\frac{n}{2}+1\right)^n}{2}$$

- Standard Deviation
  In statistics, the standard deviation (SD) is a widely used measure of variability. It depicts the degree of deviation from the average (mean). A low standard deviation (SD) suggests that the data points are near to themean.

$$\text{Standard Deviation } (\sigma) = \sqrt{\Sigma}\ (xi - \mu)^2/N$$

Predictive Analytics

Predictive analytics is a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events.2 Unlike other BI technologies, predictive analytics is forward-looking, using past events to anticipate the future (Eckerson, 2007).

- *Random Forest*
  One of the most efficient categorization algorithms is the Random Forest (RF) algorithm. It attracts scholars from several fields due to its inherent multidisciplinary nature. The goal of this research is to see how well the RF method performs with multispectral satellite photos of various spatial resolutions and scene features (Lin, 2017).
- There are some advantages in using Random Forest:
  1. Random forest can solve both classification and regression issues and provide reasonable estimation in both cases.
  2. One of the most appealing features of Random Forest is its ability to handle big data sets with higher dimensionality. It is one of the dimensionality reduction methods since it can handle hundreds of input variables and identify the most significant variables. In addition, the model outputs the importance of each variable, which is a very useful feature.
  3. It has a good strategy for predicting missing data and keeps its accuracy even when a lot of data is missing.
- Disadvantages using Random Forest Algorithm:
  1. It performs well in classification, but not in regression, as it does not provide precise continuous nature prediction. In the case of regression, it cannot forecast beyond the range of the training data, and it is possible that they would overfit noisy data sets.
  2. For statistical modelers, random forest can feel like a black box method because we have very little control over what the model does. At the very least, you can experiment with different parameters and random seeds.
- *Naive Bayes*
  One of the most often used data mining techniques is Nave Bayes. Its efficiency stems from the assumption of attribute independence, which, in many real-world data sets, may be violated.

Jurnal AC USI AccUSI Journal Of Accounting USI
VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

Many approaches have been used to alleviate the assumption, with attribute selection being one of the most important. Traditional approaches to attribute selection in naïve Bayes, on the other hand, have a high computational overhead (Chen, 2020).

A naïve Bayes classifier is an algorithm that classifies things using Bayes' theorem. Naïve Bayes classifiers are based on the assumption of substantial, or naïve, independence between data point properties. Spam filters, text analysis, and medical diagnosis are all common applications of naïve Bayes classifiers.

A classification system utilizing probability and statistical approaches devised by an English scientist named Thomas Bayes that predicts future possibilities based on previous experience is known as Naïve Bayes (Saleh, 2015).

The training and classification processes are the two stages of the Naïve Bayes classification method. Naïve Bayes is a classification method that uses a limited amount of training data to determine the predicted parameters required for classification. In most complex real - world scenarios, Naïve Bayes also outperforms expectations (Saleh, 2015).

- Some of Naive Bayes Advantages are:
  1. This algorithm is fast and can help you save a lot of time.
  2. For multi-class prediction issues, Naive Bayes is a good choice.
  3. If the premise of feature independence remains true, it can outperform other models while using far less training data.
  4. Categorical input variables are more suited to Naive Bayes than numerical input variables.
- The Disadvantages of Naive Bayes are:
  1. In Naive Bayes, all predictors (or features) are assumed to be independent, which is rarely the case in real life. This limits the algorithm's usability in real-world scenarios.
  2. The 'zero-frequency problem' occurs when an algorithm assigns zero probability to a categorical variable whose category in the test data set was not present in the training dataset. To get over this problem, you should employ a smoothing approach.
  3. You shouldn't take its probability outputs seriously because its estimations can be off in some instances.

- *Decision Tree*

  Decision Tree algorithm is designed for identifying massive data sets and streaming data and runs in a distributed setting. It has been proved to be as accurate as a traditional decision tree classifier while also being scalable for streaming data processing on numerous CPUs. These conclusions are backed up by a thorough examination of the algorithm's accuracy (Ben - Haim. 2010).
  - The Advantages in Using Decision tree algorithm:
    1. Are easy to comprehend and interpret. After a quick explanation, people can understand decision tree models.
    2. Have value even if there isn't a lot of hard data. Experts' descriptions of a situation (its choices, probabilities, and costs) and their preferences for outcomes might yield important insights.
    3. Assist in determining the worst, best, and expected outcomes for various circumstances.
  - Disadvantages using decision tree:
    1. They're unstable, which means that a slight change in the data can result in a significant change in the structure of the best decision tree.
    2. They are frequently insufficiently accurate. With same data, several alternative predictors do better. A random forest of decision trees can be used to replace a single decision tree, however a random forest is not as straightforward to comprehend as a single decision tree.
    3. Information gain in decision trees is biased in favor of qualities with more levels when data includes categorical variables with differing numbers of levels.

Prescriptive Analytics

Prescriptive Analysis combines insights from previous social media analysis results to determine which action to take in a current problem or decision, using a variety of statistical

VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

techniques such as predictive analytical modelling, which is machine learning and data mining analyzing to make predictions about future or unknown events such as consumers' interest in total interactions based on its type. Prescriptive analytics, on the other hand, tells users which activities are most likely to yield the greatest benefit when the expected event occurs. The algorithm can produce a good response based on the individual's genetic makeup. Prescriptive analytics is built on AI approaches like machine learning, which refers to the ability of unassisted computer systems to perceive, adapt, and learn (Fiet, 2007).

Facebook

Facebook is a two - syllable word derived from the words face and bookbut that does not mean the face in a book or a book is in the form of a face. So, to put it simply, the word Facebook means advance book. Facebook, on the other hand, is a social networking site where users can not only display their looks and information, but also communicate with other users from all over the world.

In 2004, Facebook was created exclusively for Harvard University students. The following year, Facebook launched a service for schoolchildren to join. Facebook only opened membership to everyone in 2006, so people can now access it from anywhere and at any time. Facebook is a platform that people can express their feeling in it with post picture, video, and link, or giving reaction into another people picture or videos all around the world.

Data Mining

Data mining is a series of processes to explore added value in the form of information that has not been known manually from a database (Vulandari, 2017).

According to Suyanto data mining is an activity that includes the collection and use of historical data to find regularities, patterns, or relationships in large data sets (Suyanto, 2017). The output of this data mining can be used to improve decision making in the future.

Meanwhile, according to (Muzakir and Wulandari, 2016) Data Mining is a process of looking for an interesting pattern or information in selected data using certain techniques or methods.

Cross Validation

Cross Validation (CV) is a statistical approach for evaluating the performance of a model or algorithm in which the data is divided into two parts, which is learning process data and validation or evaluation data.

Cross validation is the process of separating a dataset into two portions, one of which is used as training data and the other as testing data. Some studies split the data into ten sections, with 90 percent being utilized for training and the remaining ten for assessment. This procedure is performed up to ten times until all data records have been incorporated into the testing data. 10 fold cross validation is another name for this procedure. 10 folds cross validation is widely used by researchers because it is proven to produce a more stable algorithm performance.

Rapid Miner

Rapid Miner is a data science software platform built by the same-named firm that combines data preparation, machine learning, deep learning, text mining, and predictive analytics into a single environment. It is utilized for commercial and business purposes.

Rapid Miner is a GUI (Graphical User Interface) display software produced by Dr. Markus Hofmann of the Blanchardstown Institute of Technology and Ralf Klinkenberg of rapid-i.com to make it easier for users to use. Rapid Miner is open source software that was produced with Java programs and is licensed under the GNU Public License. It can be executed on any operating system. Rapid Miner does not require any particular coding abilities because all of the necessary features are already included. Rapid Miner is a program that focuses on data mining. Bayesian Models, Modeling, Tree Induction, Neural Networks, and other models are among the many and comprehensive models available.

Jurnal AC USI AccUSI Journal Of Accounting USI
VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

III. DATA AND METHODOLOGY

Table 1. Data Description

Variable	Descriptions	Role
Type	The type of the post (Photos,Video,Link, or status)	Independent Variable
Page total like	The total of like in one post	Independent Variable
Total Interactions	Total of Intereactions in that post (the sum of total like, comment, and share)	Dependent Variable
Category	Manual content characterization: action (special offers and contests), product (direct advertisement, explicit brand content), and inspiration (non-explicit brand related content).	Independent Variable
Post Month	Month the post was published (January, February, March, …, December).	Independent Variable
Post Hour	Hour the post was published (0, 1, 2, 3, 4, …, 23).	Independent Variable
Post Weekday	Weekday the post was published (Sunday, Monday, …, Saturday).	Independent Variable
Paid	If the company paid to Facebook for advertising (yes, no).	Independent Variable
Lifetime post total reach	he number of people who saw a page post (unique users). Impressions are the number of times a post from a page is displayed, whether the post is clicked or not. People may see multiple	Independent Variable
Lifetime post total impressions	Impressions of the same post. For example, someone might see a Page update in News Feed once, and then a second time if a friend shares it.	Independent Variable
Lifetime engaged users	The number of people who clicked anywhere in a post (unique users).	Independent Variable
Lifetime post consumers	The number of people who clicked anywhere in a post.	Independent Variable
Lifetime post consumptions	The number of clicks anywhere in a post.	Independent Variable
Lifetime post impressions by people who have liked a page	Total number of impressions just from people who have liked a page	Independent Variable
Lifetime post reach by people who like a page	The number of people who saw a page post because they have liked that page (unique users).	Independent Variable
Lifetime people who have liked a page and engaged with a post	The number of people who have liked a Page and clicked anywhere in a post (Unique users).	Independent Variable
Comments	Number of comments on the publication	Independent Variable
Share	Number of times the publication was shared	Independent Variable
Likes	Number of likes on the publication	Independent Variable

Data collection

The data set was retrieved from UCI Machine Learning Repository that currently maintains 622 data sets as a service to the machine learning community which consists of data science, engineering, business, and games. To avoid any trends for a specific campaign, special day, user profile, or time, the dataset is structured so that each session is a unique user over the course of a year.

Jurnal AC USI AccUSI Journal Of Accounting USI
VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

Rapid Miner Studio Educational 9.10.001 was used in this study. We compare three algorithms in this study to discover the best method for dealing with number interaction in facebook data that is predicting social media performance metrics and evaluation of the impact on brand building prediction datasets by comparing: Naïve Bayes, Random Forest, and Decision Tree.

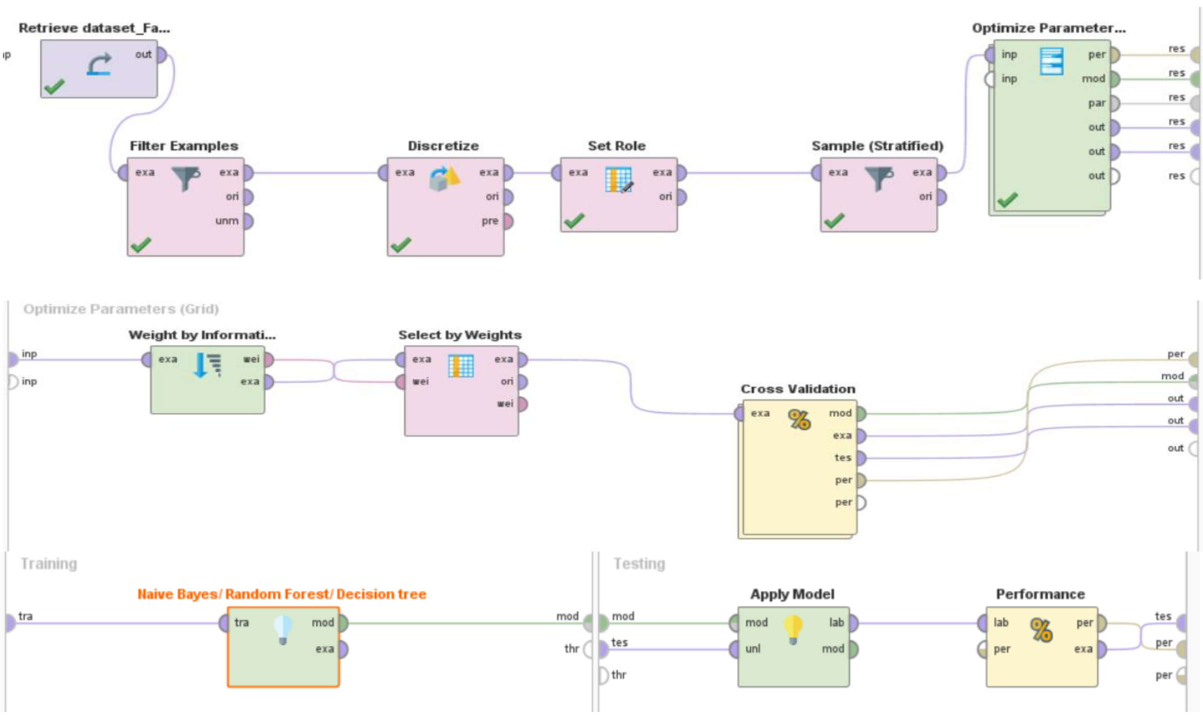


Figure 1
Classification Task Using Random Forest, Naïve Bayes, and Decision tree Algorithm

- ***Random Forest.***

We utilize the Optimize Parameter (Grid) operator to run subprocesses for all of the parameter value combinations we've chosen, and then send the best results to the parameter set port. The Random Forest Operator is the operator that will be optimized. Number of Trees is the optimal Random Forest parameter, with a Min Range of 0.05, a Max Range of 10, and a Step 20 of 20. Random Forest.apply pruning and Random Forest.minimal gain are two examples of selected parameters.

To avoid the problem of overfitting, a 10 - fold cross - validation strategy was adopted. The learner is validated using the cross validation operator. The nested operator Cross Validation is a nested operator. This operator is divided into two sub - processes: training and testing. The trained model is then put to the test. During the second phase, the performance is evaluated (Arunadevi, 2018)

- ***Naïve Bayes***

Sentiment Analysis Of Social Media Datasets That Works Extraction of subjective information from textual data is required when using Naïve Bayesian Classification. Based on its knowledge of word polarity, a normal human may quickly interpret the sentiment of a document written in natural language. (Gurkhe, 2014). But, the goal of this project is to train a machine to extract classes (low interaction, medium engagement, and high interaction) from different types of Facebook posts.

Because we are comparing between the algorithm, then we also use optimize parameters (Grid) operator in this naïve bayes algorithm, with the laplace_correction, and with 10 folds cross validation with automatic sampling type, because it is proven to produce a more stable algorithm performance.

Jurnal AC USI AccUSI Journal Of Accounting USI

VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

- ***Decision Tree***

In this decision tree, we use Decision Tree.apply_pruning, and Decision Tree.minimal_gain with cross validation 10 folds. All of these three algorithm is using Sample (Stratified) to make the process of the data quicker due to there are 500 data in the dataset.

Prescriptive Analytics

To achieve a goal, prescriptive analytics generates specific action recommendations. To put it another way, they create a link between pure analysis and actual optimization. We see two sorts of prescriptive analytics systems in general: (1) recommender systems that use data mining approaches, and (2) expert systems that use rule-based, case-based, and model-based reasoning techniques. (Gröger, 2014).

In this Prescriptive Analytics, we use firstly use Filter example to filter the missing value into is not missing, because some attributes in our data set have missing values, which was missing in the paid, like, and share atrributes.

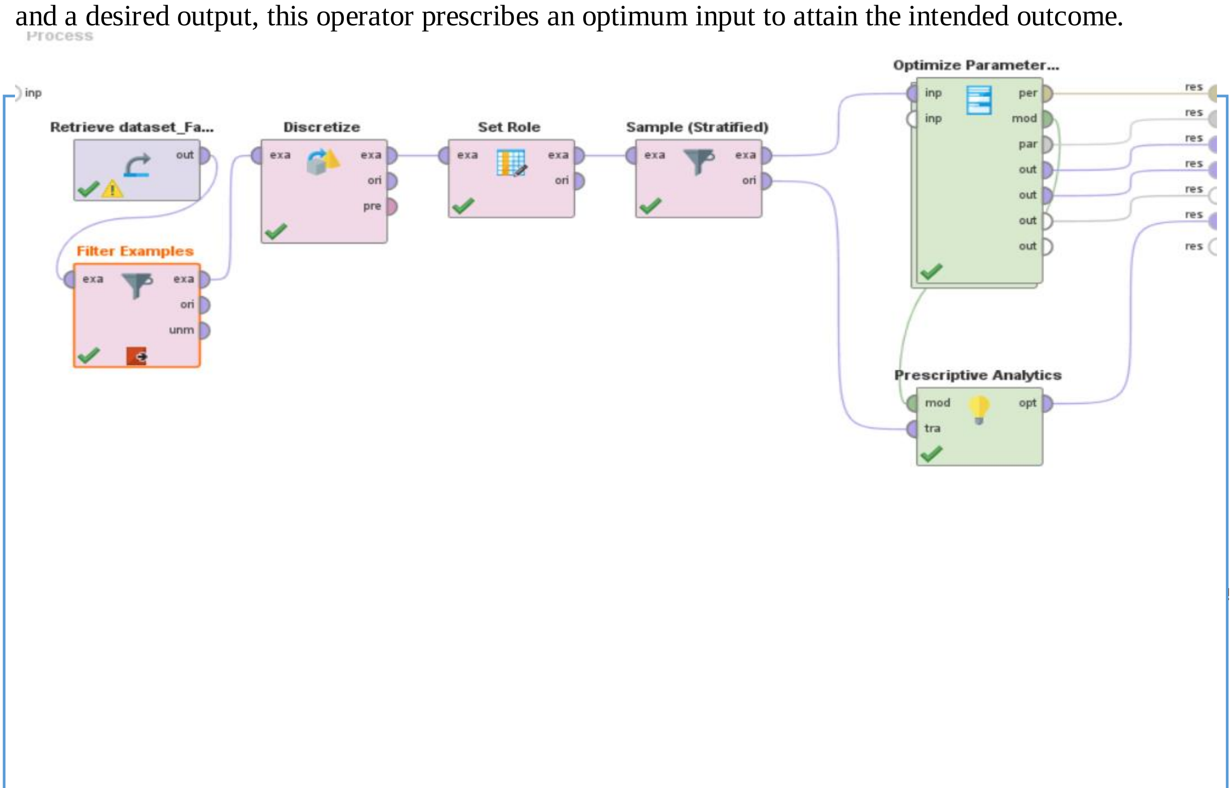
To make the classes (Low, Medium, and High Interactions) we need discretize to grouping the data, so that the dataset is easier to read and understand.After using Discretize, we need an attributes to become label in order to be able to process data to the next stage and the attributes that we choose to become label is Total Interactions with the "set role" attributes.

After finish setting the set role, we continue to add sample (stratisfied). From an ExampleSet, this operator builds a stratified sample. Stratified sampling creates random subgroups and assures that the subsets' class distribution matches that of the entire ExampleSet. This operator can't be used with data sets that don't have a label or that have a numerical label. The sample size can be chosen both in absolute and relative terms.

The stratified sampling creates random subgroups and ensures that the subsets' class distribution matches that of the entire ExampleSet. In the case of a binominal classification, for example, stratified sampling creates random subsets with about equal proportions of the two values of class labels in each subset.

To optimize the data, we use optimize parameters which inside it, we add Cross Validation with 5 number of folds and stratified sampling type. In the training side, we use naive bayes algorithm, and for the Testing side, we use Apply model, and Performance.Cross validation is the process of separating a dataset into two portions, one of which is used as training data and the other as testing data, so that the training side is using Naive Bayes algorithm, and the testing side is using Apply Model and Performance.

Because it is prescriptive analytics model, then it needs to use prescriptive analytics operator. A model is used in predictive modeling to predict an outcome given an input. Starting with a model and a desired output, this operator prescribes an optimum input to attain the intended outcome.



VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

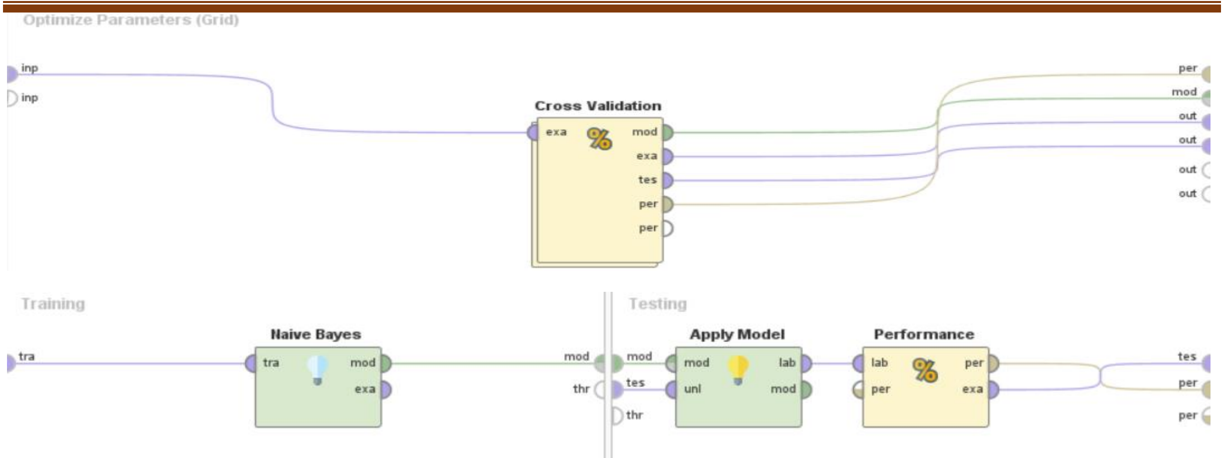
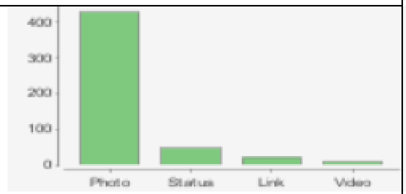


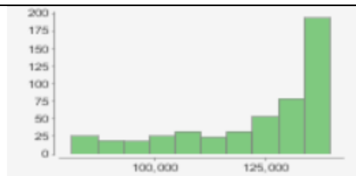
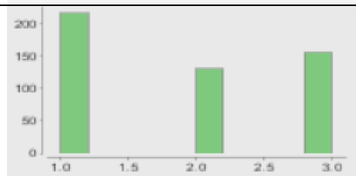
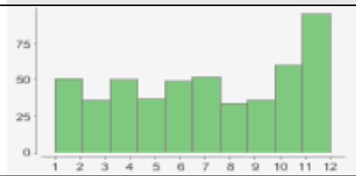
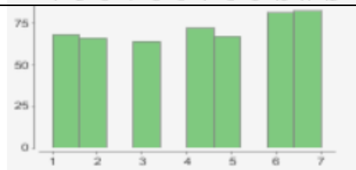
Figure 2
Classification Task Using Prescriptive Optimization

IV. RESULT AND DISCUSSION

a. Descriptive Analytics:

Table 2 Descriptive Analytics Table

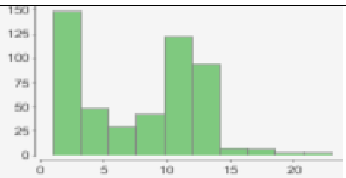
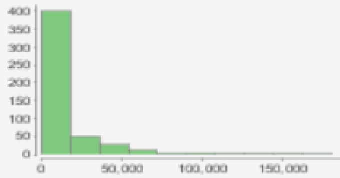
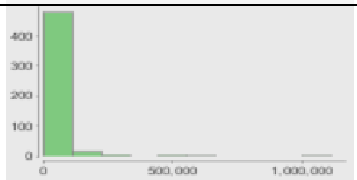
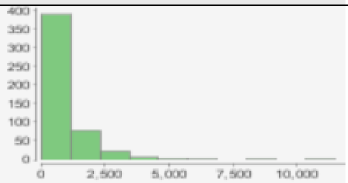
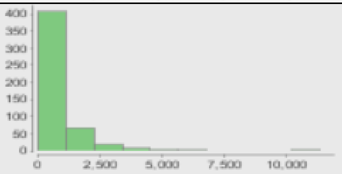
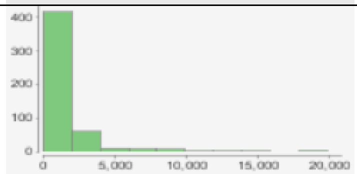
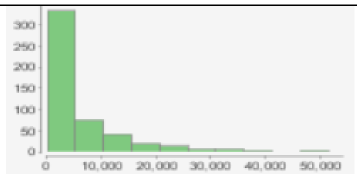
Name	Type	Least	Most	Values	Statistics
Type	Nominal	Video(7)	Photo(426)	Photos(426), Status(45), Link(22), Video(7)	

Name	Type	Min	Max	Average	Deviation	Statistics
Page total likes	Integer	81370.0	139441.0	123194.176	16272.813 214464477	
Category	Integer	1	3	1.88	0.8526746 790050889	
Post Month	Integer	1	12	7.038	3.3079360 455975397	
Post Weekday	Integer	1	7	4.15	2.0307012 323308324	

Jurnal AC USI AccUSI Journal Of Accounting USI

VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

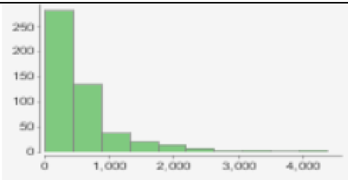
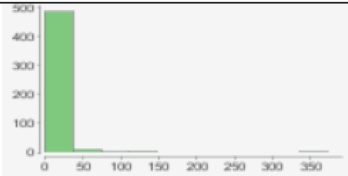
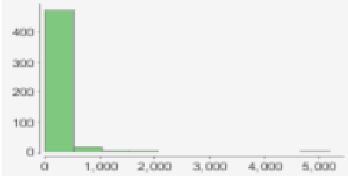
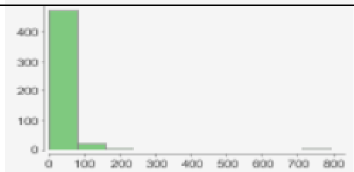
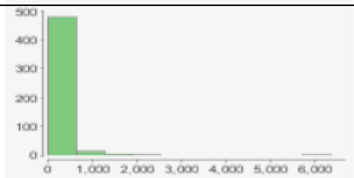
Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

Name	Type	Min	Max	Average	Deviation	Statistics
Post Hour	Integer	1	23	7.84	4.3685889 18435395	
Paid	Integer	0	1	0.278557114 2284569	0.4487388 868806484	
Lifetime Post Total Reach	Integer	238.0	180480.0	13903.36	22740.787 889562573	
Lifetime PostTotal Impressi ons	Integer	570.0	1110282. 0	29585.948	76803.246 66965121	
Lifetime Engaged Users	Integer	9.0	11452.0	920.344	985.01663 60069362	
Lifetime Post Consume rs	Integer	9.0	11328.0	798.772	882.50501 31372763	
Lifetime Post Consump tions	Integer	9.0	19779.0	1415.13	2000.5941 184441942	
Lifetime Post Impressi ons by people who have liked your Page	Integer	567.0	1107833. 0	16766.376	59791.023 73072534	
Lifetime Post reach by people who like	Integer	236.0	51456.0	6585.488	7682.0094 05283343	

Jurnal AC USI AccUSI Journal Of Accounting USI

VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

Name	Type	Min	Max	Average	Deviation	Statistics
your Page						
Lifetime People who have liked your Page and engaged with your post	Integer	9.0	4376.0	609.986	612.72561 81769623	
comment	Integer	0	372.0	7.482	21.180909 75089953	
Like	Integer	0	5172.0	177.9458917 8356713	323.39874 15906344	
Share	Integer	0	790.0	27.26612903 2258064	42.613292 15892268	
Total Interactio ns	Integer	0	6334.0	212.12	380.23311 80316835	

Above this descriptive Analytics, we can see that the total interaction in the post is reach 6334 (sum of likes, share, and comments) which the total interactions the most is got from the photo types of post, because we can see that the photo value (426) is significantly higher than the others (Status (45), Link (22), and the Video (7)). In the next step, with the predictive and prescriptive analytics can determine the prediction of total interactions in the future.These descriptive analytics are easier to use because they don't require any training, but the data processing is more complicated and the findings are less thorough.

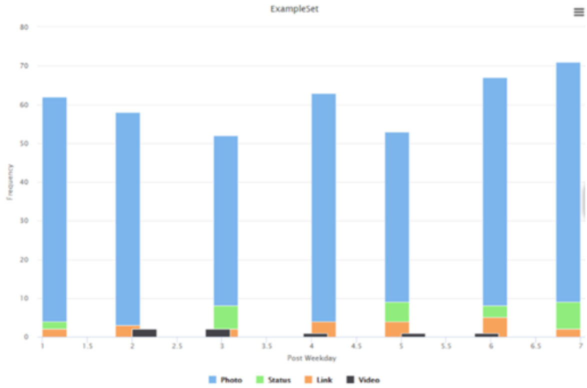


Figure 3. Statistics of Type of Post in weekdays

For the Post Weekday graphic show that from Monday until Sunday the type of Photo was the highest one for the total interaction with each result, above 60 at Monday, above 50 almost to 60 at Tuesday, above 50 at Wednesday, above 60 at Thursday, above 50 at Friday, above 60 at Saturday and 70 at Sunday. So we can conclude that photo have the most interaction in the Sunday.

Next for the Status Type the Post Weekday that the highest is at Friday, followed by Sunday, Wednesday, Saturday and Monday for the lowest one.

The Link Type show result with Saturday is the highest followed by, Thursday, Friday, Tuesday, Monday, Wednesday and the lowest is Sunday.

And for the Video Type Tuesday and Wednesday was the highest one followed by Thursday, Friday and Saturday was the lowest one.

b. Predictive Analytics

Naive Bayes

Table View Plot View

accuracy: 87.24% +/- 6.41% (micro average: 87.21%)

	true Low interaction	true Medium interaction	true High interaction	class precision
pred. Low interaction	108	22	0	83.08%
pred. Medium interaction	12	135	4	89.40%
pred. High interaction	0	0	16	100.00%
class recall	90.00%	85.99%	80.00%	

Figure 4. Confusion Matrix of Naive Bayes

The result from Naïve Bayes show that the accuracy is 87.24% with the pred. low interaction is 108 with the class precision 83.08%, pred. Medium interaction is 135 with class precision 89.40% and pred. High interaction is 16 with the class precision 100%. And for the each class recall is show with the result 90% (true Low interaction), 85.99% (true Medium interaction), and 80% (true High interaction).

Random Forest

accuracy: 90.35% +/- 4.31% (micro average: 90.37%)

	true Low interaction	true Medium interaction	true High interaction	class precision
pred. Low interaction	121	24	0	83.45%
pred. Medium interaction	0	134	5	96.40%
pred. High interaction	0	0	17	100.00%
class recall	100.00%	84.81%	77.27%	

Figure 5. Confusion Matrix of Random Forest

From Random Forest , the accuracy was the highest one with the 90.35% accuracy. The result from we using Random Forest is that the pred. Low interaction is 121 with class precision 83.45% and the class recall true Low interaction is 100%, pred. Medium interaction is 134 with class precision 96.4% and the class recall for the true Medium interaction is 84.81% , and pred. High interaction is 17 , the class precision is 100% and the class recall for the true High interaction is 77.27%.

Decision tree:

accuracy: 90.05% +/- 9.25% (micro average: 90.03%)

	true Low interaction	true Medium interaction	true High interaction	class precision
pred. Low interaction	120	24	0	83.33%
pred. Medium interaction	1	134	5	95.71%
pred. High interaction	0	0	17	100.00%
class recall	99.17%	84.81%	77.27%	

Figure 6. Confusion Matrix of Decison Tree

And the last we was using Decision tree with the accuracy 90.5% with the pred. Low interaction is 120 with the class precision 83.33%, pred. Medium interaction is 134 with class precision 95.71% and pred. High interaction is 17 with class precision 100%. For the each class recall result is 99.17% for the true Low interaction, 84.81% for the true Medium interaction and 77.27% for the High interaction.

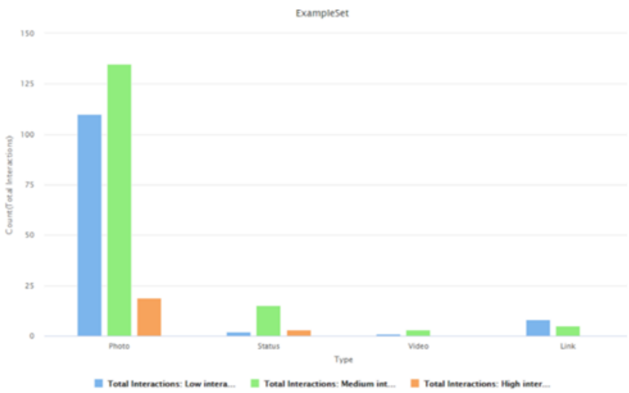


Figure 7. Statistics of Total interactions by Type of Post

The graphic show that the most highest of the Total Interactions photo is Medium Interaction that over than 125, the second highest Total Interaction is the Low Interaction that over 100 and the lowest Total Interaction is High interaction with under 25.

For the Total Interactions of status the highest one was the Medium Interaction , and the lowest is High interaction and Low interaction, but the Low interaction is more lower than the High interaction, so we can say that the Medium interaction was the highest one for the status Total Interactions, the second was High interaction and the lowest was Low Interaction.

Jurnal AC USI AccUSI Journal Of Accounting USI

VOLUME 5 No 1 Mei 2023
E – ISSN : 2620 – 5815
DOI : 10.36985/jia.v5i1.636

Jurnal Ilmiah AccUsi http: //jurnal.usi.ac.id/index.php/ACCUSI

The video type of the total interactions that more highest one is the Medium Interaction and the lowest one was the Low Interaction.

And for the type Link , the Total Interactions that the highest one is Low interaction and Medium Interactions is the lowest one.

Prescriptive Analytics:

- From the prescriptive analysis, we can see the result as shown below:

Row No.	prediction(T...	confidence(...	confidence(...	confidence(...	Category	Post Weekd...	Post Month	like
1	High Interacti...	0	0	1	2	5	9	424

Page total li...	Lifetime Peo...	Paid	Post Hour	Lifetime Eng...	Lifetime Post Impressions by people who have liked your Page
116499	412	0	6	34	68553

Type	Lifetime Post Consumptions	Lifetime Post Total Impressions	Lifetime Post Total Rea...	comment	share
photo	-435	-27488	11549	16	53

Lifetime Pos...	Lifetime Post reach by people who like your Page
853	9478

This is the result of prescriptive analytics by using the process shown in the Chapter 3 Figure 2. As we stated above that the attributes that we analyze for prediction is the total interactions , and the type that we analyze is Photo. As we can see from the result stated above, it shows that the total interaction prediction is HIGH with the confidence level of high has the highest value compared to the low and mid, which is 1while the low and medium values are 0. The category that get high interaction is Product (direct advertisement, explicit brand content) with the photo type of post. Meanwhile, the post hour is at 6.AM, with the page total like is 116499, and the lifetime enganged users is 34.

V. CONCLUSION

This study which uses the business analytics concept, is able topredict how users interact on Facebook with several post categories such as action, product, and inspiration categories, and post types such as photos, statuses, videos, and links. The total interaction on Facebook has 20 variables. We also use the information gain ratio to apply dimensionality reduction. Naive Bayes, Random Forest, and Decision Tree are the algorithms we use. The results of the comparison of the three algorithms that we use show that Random Forest is the algorithm with the highest accuracy, which is 90.35 %. Besides, 3 business analytic domainsare utilized, such as descriptive - analytic, predictive analytic, and perspective analytic. The Algorithms used in predictive analytics include Random forest, Naive Boyes, and Decision trees.

By using the business analytic concept that utilizing *rapid miner* software, it providesrobust and cost-friendly analytical computation processesin predicting total interactions on Facebook and also determines predictions about how users interact on Facebook with several post categories such as action, product, and inspiration categories.

This can make it easier for an analyst to determine whether a post is worth paying for and promoting. This data is sought for deeper observations of posts on Facebook that get a lot of interaction and loyal users. However, the weighting results showed only minor variances in value among the 20 attributes we used. More daily data could provide a broader and more complete picture. As a result, using a larger and more complicated dataset in future research may uncover further secret, useful, and hidden information in Dataset Facebook.

REFERENCES

Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next?. Business Horizons, 57(5), 565-570.

Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2010, March). Naive Bayes variants in classification learning. In 2010 International Conference on Information Retrieval & Knowledge Management (CAMP) (pp. 276-281). IEEE.

Arunadevi, D. J., Ramya, S., & Raja, M. R. (2018). A study of classification algorithms using Rapidminer. International Journal of Pure and Applied Mathematics, 119(12), 15977-15988.

Basri, W. S., & Siam, M. R. (2017). Maximizing the social media potential for small businesses and startups: A conceptual study. International Journal of Economic Perspectives, 11(2), 241-245.

Ben-Haim, Y., & Tom-Tov, E. (2010). A Streaming Parallel Decision Tree Algorithm. Journal of Machine Learning Research, 11(2).

Bianchi, C., & Andrews, L. (2015). Investigating marketing managers' perspectives on social media in Chile. Journal of Business Research, 68(12), 2552–2559.

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. Knowledge-Based Systems, 192, 105361.

Eckerson, W. W. (2007). Predictive analytics. Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report, 1, 1-36.

Edosomwan, S., Prakasan, S. K., Kouame, D.,Watson, J., & Seymour, T. (2011). The history of social media and its impact on business. Journal of Applied Management and Entrepreneurship, 16(3), 79–91.

Fiet, J. O. (2007). A prescriptive analysis of search and discovery. Journal of Management Studies, 44(4), 592-611.

Friedman, J. H., Kohavi, R., & Yun, Y. (1996, August). Lazy decision trees. In AAAI/IAAI, Vol. 1 (pp. 717-724).

Griffiths, M., & McLean, R. (2015). Unleashing corporate communications via social media: A UK study of brand management and conversations with customers. Journal of Customer Behaviour, 14(2), 147–162

Gröger, C., Schwarz, H., &Mitschang, B. (2014, May). Prescriptive analytics for recommendation-based business process optimization. In International Conference on Business Information Systems (pp. 25-37). Springer, Cham.

Gurkhe, D., Pal, N., & Bhatia, R. (2014). Effective sentiment analysis of social media datasets using Naïve Bayesian classification. International Journal of Computer Applications, 975(8887), 99.

Habibi, M. R., Laroche, M., & Richard, M. O. (2014). The roles of brand community and community engagement in building brand trust on social media. Computers in Human Behavior, 37, 152–161.

Kemp, S. E., Ng, M., Hollowood, T., &Hort, J. (2018). Introduction to Descriptive Analysis. Descriptive Analysis in Sensory Evaluation, 1-39.

Kraus, S., Kanbach, D. K., Krysta, P. M., Steinhoff, M. M., &Tomini, N. (2022). Facebook and the creation of the metaverse: Radical business model innovation or incremental transformation. International Journal of Entrepreneurial Behavior& Research.

Lariscy, R.W., Avery, E. J., Sweetser, K. D., & Howes, P. (2009). Monitoring public opinion in cyberspace: How corporate public relations is facing the challenge. Public Relations Journal, 3(4), 1–17.

Lin, W., Wu, Z., Lin, L., Wen, A., & Li, J. (2017). An ensemble random forest algorithm for insurance big data analysis. Ieee access, 5, 16568-16575.

Trainor, K. J., Andzulis, J. M., Rapp, A., & Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. Journal of Business Research, 67(6), 1201–1208.1016/j.jbusres.2013.05.002.

Turban, E., Sharda, R., Delen, D., & Efraim, T. (2011). Decision support and business intelligence systems (9th ed.). Pearson.

Xanthidis, D., & Nicholas, D. (2004, December). Evaluating internet usage and ecommerce growth in Greece. In Aslib Proceedings. Emerald Group Publishing Limited.