INVESTIGATING MARKETPLACE CUSTOMER SATISFACTION: DATA MINING APPROACH

Suryaneta Suryaneta1), Achmad Gus Fahmi2), Iwan Syahjoko Saputra3), Kiki Yuli Handayani4)
1 Program Studi Rekayasa Kosmetik, Jurusan Teknologi Produksi dan Industri, Institut Teknologi Sumatera
email: suryaneta@km.tera.ac.id
2 Program Studi Rekayasa Kosmetik, Jurusan Teknologi Produksi dan Industri, Institut Teknologi Sumatera
email: achmad.fahmi@km.tera.ac.id
3 Program Studi Rekayasa Kosmetik, Jurusan Teknologi Produksi dan Industri, Institut Teknologi Sumatera
email: iwan.saputra@km.tera.ac.id
4 Program Studi Rekayasa Kosmetik, Jurusan Teknologi Produksi dan Industri, Institut Teknologi Sumatera
email: kiki.handayani@km.tera.ac.id

Abstract
Customer satisfaction is perceived as business strategy’s key component and critical differentiation in a competitive marketplace. The recorded huge transactions per day in marketplaces can be used as useful information to evaluate customer satisfaction. A sophisticated method, such as data mining, is necessary to analyze this massive, multifaceted, and versatile empirical data generating accurate predictions. This research purported to investigate marketplace customer satisfaction as a reference to determine service and quality improvements. In conclusion this study draws two conclusive results: (1) The majority of marketplace consumers preferred the lead-time sensitive over price-sensitive; and (2) The Neural Net empirically showed as the most appropriate robust data mining technique among other techniques to predict marketplace customer satisfaction indicating by fittest accuracy, F score and ROC curve.

Keywords: Customer satisfaction, data mining, machine learning, marketplace, online shopping

1. INTRODUCTION
The number of transactions in online retail continues to grow in the pandemic era. A number of intriguing features have contributed to the development of a digital ecosystem that encourages people to shop online rather than in-store. Customers can order products and have them delivered to their destination without having a physical presence. Unlike physical stores, where customers can physically inspect and test products, digital services rely heavily on customer feedback via ratings and reviews. Customer satisfaction is quantifiable through their ratings of the services they receive.

Maintaining a high level of service quality from the customer's perspective is critical for any business, and a satisfying online shopping experience is a critical component of meeting customer convenience [1]. As a result, customer satisfaction must be built and maintained to the maximum extent possible through the provision of superior products and services in order to maintain relationships with customers [2]. Customers have the right to provide feedback, either positive, neutral, or negative, regarding their purchasing experience via the review rating facility. This rating is an evaluation material for the subsequent development of business by maintaining a positive response and correcting negative responses for the better. Also, customers' reviews can help sellers or decision-makers determine which products are in demand or what kind of service needs to be fixed.

Several large marketplaces have even seen nearly thousands of transactions in a single day [3]. Advanced methods as evaluation materials are required to analyze the massive, complex, and flexible empirical data in order to
extract useful information for use. Data mining approach is well known as a solution for complex data processing by providing more accurate methods for making predictions [4].

In this paper, the authors examined marketplace customer satisfaction through the lens of a machine learning algorithm, focusing on the type of satisfaction desired by marketplace customers via feature selection. This study’s results can be used by service providers to further improve service quality according to empirical data. In order to obtain more accurate predictive results and to compare them, authors employed multiple classifiers rather than a single classifier. Several classification algorithms were used in this study, including Random Forest, Decision Tree, k-NN, Naïve Bayes, Rule Induction, Neural Net, Logistic Regression and Support Vector Machine. The purpose of this study is to investigate customer satisfaction in the marketplace in order to use it as a benchmark for service enhancements, quality enhancements, and the development of more effective marketing strategies.

LITERATURE REVIEW

Marketplace Customer Satisfaction

The development and advancement of digital technology are so fast that it gives rise to behavior changes in most people. It can be seen from the shopping habits that are currently fonder of shopping online than before being accustomed to shopping directly at the nearest market, shop or mall. The main reason for this shift in behavior from offline to online is based on the marketplace’s convenience. The marketplaces are a fundamental platform that helps put together a large number of suppliers and buyers. They offer a link between various participants and provide necessary conditions and guarantee.

Additionally, customer satisfaction is an emotional response to the discrepancy between what customers receive and what they desire. Several research pieces have stressed the assessment of customer satisfaction in online shopping from different standpoints by conducting surveys or interviews. For example, a research by Al Karim in 2013 conducted a customer satisfaction survey in the marketplace and discovered that the term satisfaction encompasses time savings, information availability, ease of use, web surfing, less shopping stress, lower costs, and purchasing enjoyment [5]. In comparison, significant barriers to online shopping include online payment security, individual privacy and trust, vague warranties and return policies, and a lack of personalized customer service.

Without surveys or interviews, marketplace customer satisfaction assessment is by providing online consumer review and rating. According to Mo and Fan, online customer reviews are reviews provided by consumers related to evaluating a product or service on various aspects [6]. Meanwhile, consumer rating is the same thing as a review, but consumers’ opinion is an assessment of customers on the preferences of a product or services for their experience in the form of a determined scale.

Usually, the marketplace’s rating is in the form of stars, where more stars show a better satisfaction value [7]. In the rating scale, the consumer will determine from one to five stars. If the stars reach four or five stars, the goods or services on the marketplace match expectations. If the star given by the consumer is one or two stars, something is must be evaluated because the customer dissatisfied. Moreover, for three stars, it’s neutral between satisfy and not satisfy. In this research, we categorized rating scale into the binary group, which is three, four, and five rating stars as satisfying and, on the other hand, one and two rating stars as dissatisfaction.

Machine Learning Algorithms

Data mining relies heavily on classification. A machine learning algorithm is used to train and test the attributes in the dataset
to predict a categorical attribute, known as the class label, which identifies the class target. Model machine learning algorithms are evaluated by accurately predicting. There are several supervised machine learning algorithms that use in this research. Prediction for a given dataset is what matters most. Eight classifiers are Decision Tree, Random Forest, Neural Net, Deep Learning, Naïve Bayes, k-NN, Logistics Regression and Rule Induction.

2. RESEARCH METHODS

Data Source
The dataset was extracted from Kaggle dataset (Olist, a Brazilian marketplace that offers e-commerce services) that contains 100k orders from September 2016 to Augustus 2018, which is at the beginning of operating. The dataset features were divided into multiple dimensions’ datasets: order status, customer information, seller information, payment and shipping information, product attributes, review customers, and geolocation.

Data Preparation
For data preparation, all the multiple datasets were combined into a single dataset. The combination dataset using the relationship of attribute id provided in the Figure 1. This study also calculated the direct distance between geolocation customer centroid zip code coordinate and seller centroid zip code coordinate. The Great Circle formula [8] was applied for conversion geolocation coordinate between customer centroid zip code coordinate and seller centroid zip code coordinate:

$$distance = a \cos^{-1} \left[ \cos \delta_1 \cos \delta_2 \cos(\lambda_1 - \lambda_2) + \sin \delta_1 \sin \delta_2 \right]$$

(1)

a = radius of the equatorial earth (a \approx 6378206.4 meters)

δ1 = latitude of customer centroid zip code (in radians)

δ2 = latitude of seller centroid zip code (in radians)

λ1 = longitude of customer centroid zip code (in radians)

λ2 = longitude of seller centroid zip code (in radians)

The next step after getting the new attributes was data cleaning to ensure the correct, consistent, and usable data. For example, by removing duplicate data that appeared repetition for the same order_id associated with product_id to hinder the rating bias. The biases for the order_id, which has more than 1 product ordered, but the variety of general product_quantity ordered is not so large (min=1, mean=1.16, max=21, SD=0.638). The percentage of bias as:

$$\frac{N_{\text{example}} - N_{\text{order_id}}}{N_{\text{example}}} = \frac{100281 - 96614}{100281} = 3.65\%$$

(2)

Figure 1. The schema of multiple datasets

After data cleansing, the given datasets were 100,281 examples with 1 label attribute and 37 regular attributes (21 numerical and 16 categorical) with the label attribute contains 76,527 respond positive (76.31%), 8,428 respond neutral (8.41%), and 15,326 respond negative (15.28%). Then, the given datasets run for correlation matrix. Several attributes were found have insignificant relationship with all other attributes. Table 1 shows the final datasets.

Table 1. The final attributes of the dataset

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute type</th>
<th>Missing value</th>
<th>Replace missing with</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>label</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>price</td>
<td>price</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>price_gap</td>
<td>price_gap</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>product_quantity</td>
<td>product_quantity</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>shipping_cost</td>
<td>shipping_cost</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>shipping_cost_gap</td>
<td>shipping_cost_gap</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>total_cost</td>
<td>total_cost</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>payment_value</td>
<td>payment_value</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>product_name_lenght</td>
<td>product_name_lenght</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>product_descr_lenght</td>
<td>product_descr_lenght</td>
<td>numerical</td>
<td>numerical</td>
</tr>
<tr>
<td>product_photos_qty</td>
<td>product_photos_qty</td>
<td>numerical</td>
<td>numerical</td>
</tr>
</tbody>
</table>
This study applied several supervised machine learning algorithms, including Neural Net and SVM that forbid for missing value. The missing value was handled by missing replacing process as shown in Table 2.

### Table 2. Missing value replacement

<table>
<thead>
<tr>
<th>order_status</th>
<th>N_value</th>
<th>% Orde_status</th>
<th>N_missing</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>edd</th>
<th>edd</th>
</tr>
</thead>
<tbody>
<tr>
<td>delivered</td>
<td>98130</td>
<td>97.9</td>
<td>8</td>
<td>-6</td>
<td>23.90</td>
<td>154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shipped</td>
<td>1093</td>
<td>1.1</td>
<td>1093</td>
<td>2</td>
<td>25.10</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>canceled</td>
<td>447</td>
<td>0.4</td>
<td>440</td>
<td>3</td>
<td>22.70</td>
<td>145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>invoiced</td>
<td>311</td>
<td>0.3</td>
<td>311</td>
<td>3</td>
<td>25.71</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>processing</td>
<td>291</td>
<td>0.3</td>
<td>291</td>
<td>-4</td>
<td>29.77</td>
<td>145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unavailable</td>
<td>7</td>
<td>0.0</td>
<td>7</td>
<td>48</td>
<td>55.86</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>approved</td>
<td>2</td>
<td>0.0</td>
<td>2</td>
<td>22</td>
<td>22.50</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Feature Selection Method

The feature selection method in this study has employed forward selection in order to eliminate superfluous features from the clickstream dataset prior. The forwarding selection procedure is evaluated and compared to the best previous procedure using these criteria from the given dataset [9]. The forward selection process begins by identifying the feature in a subset that has a high predictive ability. If the new feature subset improves, the previous best feature subset is discarded in favor of the new feature subset. This process is repeated until no additional subset of features improves the model statistically significantly.

The purpose of this study is to compare eight supervised machine learning algorithms, including Decision Tree, Random Forest, k-NN, Naive Bayes, Rule Induction, Logistic Regression, Neural Net, and Support Vector Machine, in order to develop prediction models for online shopper behavior prediction using clickstream data. These classifier algorithms primarily made use of prediction techniques [10].

The prediction accuracy and F-score performance metrics was implemented to evaluate the prediction performance of each supervised machine learning algorithm. The confusion matrix, as shown in Table 3, can be used to quantify both evaluation metrics. The average prediction accuracy level is determined by the number of data samples classified correctly by the prediction model in a given testing set. Additionally, the F-score represents a compromise between precision and recall. The higher the F-score, the more perfect the balance of precision and recall is.

### Table 3. Confusion matrix

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Not purchase</th>
<th>Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not purchase</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Purchase</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

The prediction accuracy and F-score are obtained by:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
F1\ score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{TP}{TP + \frac{2(FP + FN)}}
\]

In addition to these two widely used metrics, this study also has examined the receiver operating characteristic (ROC) graphs of model classifiers. The graphical curve has used to visualize the predictive model as the discrimination threshold varies when adjusting a score threshold. Figure 2 demonstrates the ROC curves as very useful tool for organizing, illustrating, and analyzing prediction models based on their performance providing a performance graphing technique that is beneficial in the appearance of unbalanced classes.
One of the ROC curve’s advantages is enables visualization and classifier performance regardless of the error cost class distribution. The AUC, area under the ROC curve, define as a numerical value representing the probability of the model prediction for ranking positive instances that randomly selected higher than the negative instances.

3. RESULT

Results on Feature Selection

Table 4 informs the rank of the selected feature subset for each classification algorithm used to select features.

Table 4. Selected feature subset after feature selection process among different classifiers

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>estimated delivery duration</td>
<td>product description length</td>
<td>order status</td>
<td>payment type</td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>estimated delivery duration</td>
<td>product description length</td>
<td>order status</td>
<td>payment type</td>
</tr>
<tr>
<td>k-NN</td>
<td></td>
<td>actual delivery duration</td>
<td>on-time status</td>
<td>order status</td>
<td>payment type</td>
<td>order status</td>
<td>payment type</td>
</tr>
<tr>
<td>NB</td>
<td></td>
<td>on-time status</td>
<td>order status</td>
<td>purchase day</td>
<td>order status</td>
<td>order status</td>
<td>payment type</td>
</tr>
<tr>
<td>RI</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>estimated delivery duration</td>
<td>order status</td>
<td>distance</td>
<td>product name length</td>
</tr>
<tr>
<td>LR</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>estimated delivery duration</td>
<td>order status</td>
<td>distance</td>
<td>product name length</td>
</tr>
<tr>
<td>NN</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>estimated delivery duration</td>
<td>product description length</td>
<td>product quantity</td>
<td>approved duration</td>
</tr>
<tr>
<td>SVM</td>
<td></td>
<td>on-time status</td>
<td>actual delivery duration</td>
<td>payment instalments</td>
<td>product description length</td>
<td>product quantity</td>
<td>approved duration</td>
</tr>
</tbody>
</table>

Based on the selected feature subset’s observation, we can see that “on-time status” is in the first rank of all classifiers except k-NN which places it in the second position. Meanwhile, “actual delivery duration” ranks second in all classifiers, except k-NN, which ranks in the first position. However, it did not select by SVM as a feature subset. Meanwhile, “estimated delivery duration” is ranked third in several classifiers such as DT, RF, RI, and NN. Next is “order status” placed by NB in second place and LR placed in third place. This can be indicated that “on-time status” plays the most significant role among other features in determining customer satisfaction, followed by “actual delivery duration”, “estimated delivery duration”, and “order status”.

The rest feature selected is week predictor to determine consumer satisfaction which features subset under the top four, such as “product description length”, “purchase day”, “payment instalments”, “distance”, “product quantity”, “product name length”, “approved duration”, “payment type”, and “price gap”.

The rest of the other features are not selected in the feature selection, which can be assumed that features do not affect consumer opinion in satisfaction assessments, such as price, product quantity, shipping cost, shipping cost gap, total cost, product photos quantity, product weight, product length / high / width, sequential payment, purchase weekend, delivery weekend, review respond duration and approved duration.

Results on Performance Prediction

Table 5 demonstrates the compared prediction performance among various classification algorithms. Based on the comparison of model predictions, it can be concluded that NN is the strongest classifier in accuracy at 88.26% and also an F score of 93.39% to predict customer satisfaction.

Table 5. Prediction models of different classification algorithms.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Accuracy (%)</th>
<th>F score (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>87.89</td>
<td>93.19</td>
<td>0.673</td>
</tr>
<tr>
<td>RF</td>
<td>87.94</td>
<td>93.21</td>
<td>0.684</td>
</tr>
<tr>
<td>k-NN</td>
<td>87.21</td>
<td>92.83</td>
<td>0.658</td>
</tr>
<tr>
<td>NB</td>
<td>87.57</td>
<td>92.94</td>
<td>0.675</td>
</tr>
<tr>
<td>RI</td>
<td>88.14</td>
<td>93.34</td>
<td>0.671</td>
</tr>
<tr>
<td>LR</td>
<td>87.77</td>
<td>93.08</td>
<td>0.707</td>
</tr>
<tr>
<td>NN</td>
<td>88.26</td>
<td>93.39</td>
<td>0.728</td>
</tr>
</tbody>
</table>
Results on ROC

As an additional comparison, the visualization of the ROC as in Figure 3 shows that the ROC curve of NN shows as the best predictor with AUC = 0.728 among other classifiers in predicting customer satisfaction.

<table>
<thead>
<tr>
<th></th>
<th>SVM 87.57</th>
<th>92.94</th>
<th>0.673</th>
</tr>
</thead>
</table>

Figure 3. The ROC curve result among supervised machine learning algorithms

4. CONCLUSION

Implications for theory and research

Many businesses and the marketplace have recently been struggling to draw consumers to acknowledge their preferences. Based on consumer experience, communications and behavior, businesses respond to the need of their customers.
Marketplace or online retailers need advanced technologies to integrate and collect data regarding their customers since conventional approaches to extensive data are ineffective. However, data mining seems to be a more useful strategy to save much time and minimize human errors as a powerful tool [11].

Data mining offers analytically comprehensive customer information, valuable insights into e-commerce strategy implementation and preferences, expectations, and customer satisfaction analysis. When properly handled, data mining helps the marketplace to deliver better service at the right time as it knows what consumers want.

The present study explores the customer information data by using the data mining approach in predicting marketplace customer satisfaction. Our study showed that feature selection result for “on-time status” as first consideration of customer to predict their satisfaction, followed by “actual delivery duration” and “estimated delivery duration” as a second and third place—these three features related to the delivery time and logistics performance in delivering product from sellers to customers.

Delivery speed is very significant in the framework of e-commerce. The various studies indicate that service speed is a significant criterion in ensuring that customers are satisfied with their purchases. Previous studies have shown that the delivery speed indicator applied to a firm’s ability to deliver on or before the estimated delivery date that is promised [12]. In the meantime, another study from Rosenzweig explains that delivery time quality is determined by the actual delivery time [13]. According to Weaver-Meyers and Stolt, delivery speed is essential in achieving customer satisfaction [14]. Furthermore, Rao explain that excellent delivery performance will encourage customers to buy again or even pay for much more [15]. Another advantage from the fast delivery implies that the expected delivery duration can be used as a consumer marketing strategy.

According to Griffis show delivery time with satisfaction and quantity correlated with purchase frequency [2]. Peng and Lu (2017) argue that delivery performance impacts customer transaction quantities [16]. However, when online retailers compete with physical retailers, e-commerce can suffer tremendous losses if the delivery system responds slowly [17]. Again, Rao et al. (2011) reported online shopping delivery irregularities affecting the order’s volume, purchase frequency and customer loss. In the context of business to business, Yu et al. (2013) argued that good delivery performance would have significant consequences for customer satisfaction and positively affect the business’s performance [18].

Consumers will be more satisfied with the purchase to the customer place on time and in good condition. Furthermore, if it arrives late, it could be the last purchase a customer makes. The time way has a strong prediction for customer satisfaction. We advise the market to strengthen logistics performance because customers will be satisfied if the purchased product arrives on time from the estimated delivery day. Lee and Dinwoodie (2016) suggest that logistics integration as the backbone of supply chain management can improve business performance [19].

Managerial implications

Customer satisfaction should be a priority for every business the fulfilment. Consistent delivery problems can lead to
problems in many other areas of the supply chain. Examples include delays in production due to material shortages and customer complaints. A delivery speed creates customer’s satisfaction and reduces a sense of worry for business customers. Short delivery lead times are preferred because long waiting times interfere with the assembly process. If the ordered goods’ delivery is always late, the customers can search for goods from competing sellers. Customers can change suppliers due to the disappointment of slow delivery and switch to another seller.

The consequence of fast delivery is to cause a premium price for the customers. In terms of delivery sensitivity, customers can be classified as lead-time sensitive or price-sensitive. In supply chain management, the lead time is when an order is received until it is shipped to the customer. Please note that a high lead time will cause stock inventory and affect customers who will eventually look for other products or other suppliers.

This study indicates that the tendency of majority marketplace consumers is lead-time sensitive rather than price-sensitive. This study is shown by prioritizing the on-time status, actual delivery duration and estimated delivery duration to determine consumer satisfaction and shown by not using the shipping cost gap as consideration for customer satisfaction. This finding is in line with the study conducted by Zhao et al. (2012), who revealed that given the sensitivity of customers prefer a faster delivery even with a higher shipping price. Thus, a seller may upgrade higher prices for shorter delivery lead times [20].

After running several different classification algorithms (Decision Tree, Random Forest, k-NN, Naïve Bayes, Rule Induction, Logistic Regression, Neural Net, and Support Vector Machine) and investigating the receiver operating characteristic (ROC) graphs of model classifiers to examine prediction models’ visual performance. The results show that Neural Net is the fittest accuracy and F score, and the fittest in visual the ROC curve as some robust data mining techniques to predict marketplace customer satisfaction.

5. REFERENCES


