A Predictive Data Analytics Methodology for Online Food Delivery

Mariam Al Akasheh[1,2]
[1]*United Arab Emirates University*
Al Ain, UAE
[2]*Al Ain University*
Al Ain, UAE
mariam_okasheh@yahoo.com

Nehal Eleyan[1]
[1]*United Arab Emirates University*
Al Ain, UAE
nehaleleyan@hotmail.com

Gurdal Ertek[1]
[1]*United Arab Emirates University*
Al Ain, UAE
gurdal@uaeu.ac.ae

***Abstract*— Online food deilvery (OFD) has become one of the popular and profitable e-business categories due to the rising demand for online food delivery. People are increasingly ordering food online, especially in urban areas and on college campuses. Using data from online food delivery services, one can analyze and predict the values of key performance indicators (KPIs). In the study presented in this paper, we developed a systematic methodology to analyze and predict such KPIs using various classification and regression algorithms. We found that, for the case study we analyzed, Random Forest (RF) consistently ranked as the best algorithm for regression and classification in predicting most of the KPIs. The methodology that we introduce and illustrate in the paper can be adapted and extended to similar problems, to reveal potential operational problems as well as identify the possible root causes of such problems.**

Keywords*— *online food delivery, food delivery services, machine learning, data analytics, data science pipeline

I. Introduction

E-commerce is growing around the world while the digital economy is growing and as more people are getting access to high-speed Internet. Especially following the COVID-19 pandemic, consumers are utilizing online services more frequently, including online food delivery (OFD) services [1]. OFD has emerged as a significant channel to reach customers and provide superior services in the modern era, playing the crucial role of sustaining previously brick-and-mortar food services in the digital age [2]. In addition to simplifying the ordering process for both the consumer and the restaurant, the primary benefit of this type of service is that it retains customers and improves the customer experience by providing new delivery channels [3].

Given the increasing importance of OFD, there are many research opportunities, including the development of new data analytics methodologies for analyzing data from OFD. In our study, we developed and demonstrated the applicability of such a data analytics methodology, which consists mainly of classification and regression analysis to predict the key performance indicators (KPIs) in OFD. We compared 11 classification algorithms and four regression algorithms, reporting our methodology and findings in this paper.

The paper is structured as follows: This first section introduces the problem and describes the scope of the study. Section 2 summarizes the earlier relevant work in the literature. Section 3 describes the methodology that was developed and applied throughout the study. Section 4 presents the analysis results, demonstrating the types of insights that can be obtained through the presented methodology. Section 5 provides concluding remarks.

II. Literature Review

Our work is based mainly on the earlier work by Correa et al. (2019) [4], whose source data [5] was published as an appendix to separate paper [6]. The authors in [4] analyzed data collected from an OFD service in Bogotá, Columbia, and evaluated the effect of traffic conditions on the KPIs of OFD. The authors, conducting descriptive and diagnostic analytics, focused mainly on the distribution of KPI values and the correlations between KPIs and attributes such as traffic intensity and customer comments. While the results presented by the authors provide interesting and beneficial insights, their study [4] did not include the application of predictive analytics, which was chosen as the main topic and target contribution of our paper.

Teichert et al. (2020) [7] employed text mining to analyze the experiences of customers of fast-food delivery services. The study revealed that consumers value fast meal delivery for more than just its quickness. Customers care more about how their service is handled than how their products are made or how their payments are handled. This means that OFD services must make sure that their products and services are both excellent.

Modak et al. (2019) [8] applied a logistic regression model to determine which attributes influence the customer satisfaction of OFD services. The following attributes were considered: higher discounts, a wider selection of restaurants; food quality; packaging; on-time doorstep delivery; customer service; payment options; and pricing. Better discounts, restaurant selection, food quality, and payment methods were found to be positively associated with customer satisfaction. In our study, we also apply logistic regression to characterize binary KPIs.

Using the expanded theory of planned behavior (TPB), Prasetyo et al. (2021) [9] determined the factors that influence customer satisfaction and loyalty in food delivery services during the COVID-19 pandemic. According to the findings, there are four factors that influence consumer satisfaction and loyalty. Hedonic motivation (HM) is the main factor, followed by price, information quality, and marketing.

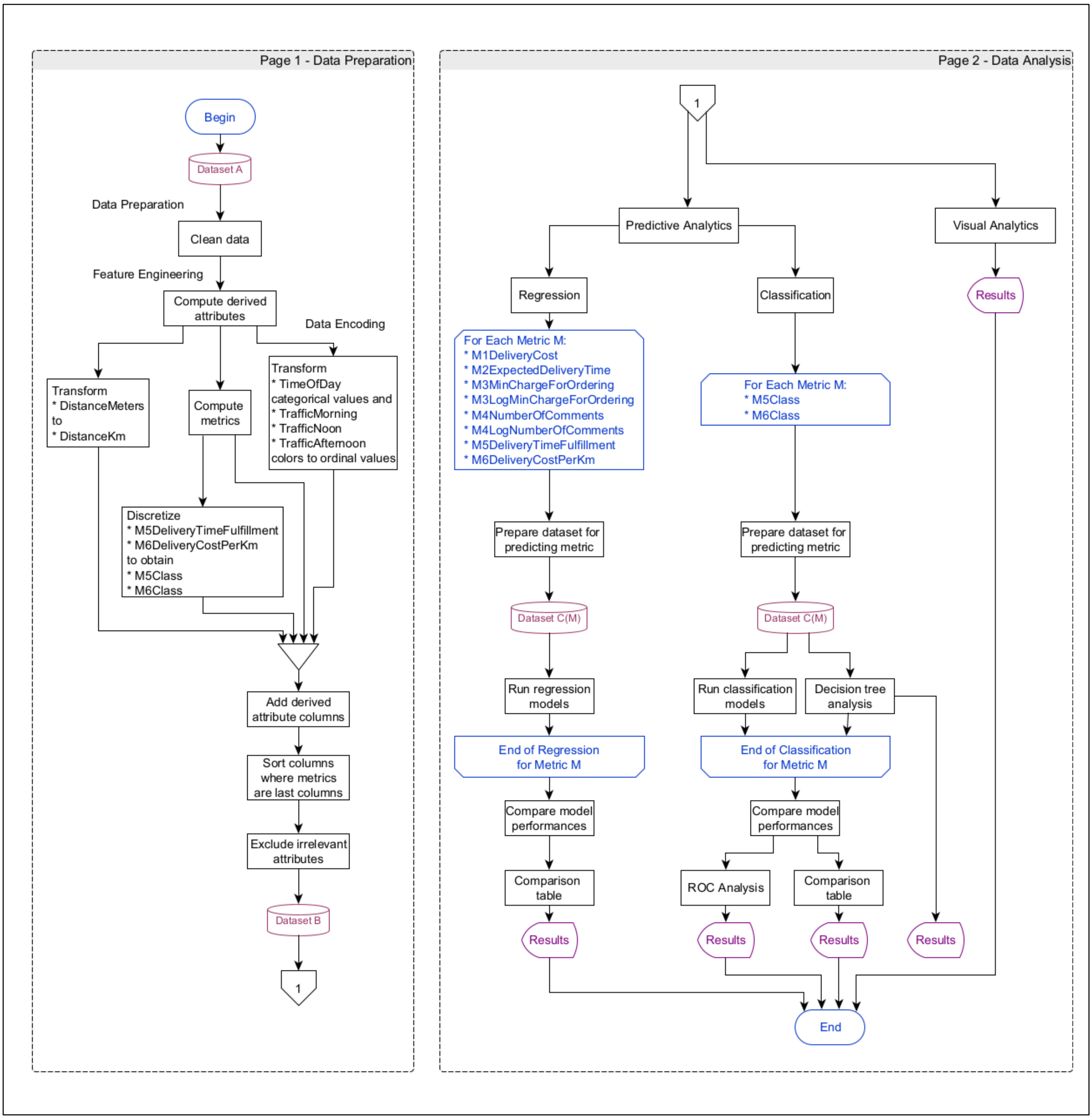

Fig. 1. Methodology developed for and applied in the presented study.

A critical issue in data driven OFD research is the measurement of performance, which requires the definition and development of the most relevant scales. To this end, Cheng et al. (2021) [10] employed qualitative and quantitative research methods and incorporated the major service elements for the online food delivery (OFD) business to develop an OFD service quality scale (OFD-SERV). The results revealed that customer satisfaction is improved by system reliability, system assurance, and system operation.

III. METHODOLOGY

The topic of this paper is the analysis of data from OFD services, using data analytics -primarily predictive analytics- techniques.

For analyzing data from any domain, the development, documentation, and application of a systematic data analytics methodology can bring many benefits. Firstly, it can enable reproducible analyses and results. Secondly, the presence of such a methodology increases the validity of the study and the reliability of the results. Thirdly, a well-documented methodology can be adopted and extended much more easily.

Table I. Features (attributes) in the constructed Dataset B, their data types, descriptions, and source.

Features (Attributes)	**Data type**	**Description**	**Source**
Transaction ID	Numeric	ID for each transaction	NewAttribute
TimeOfDay	Ordinal	time slices in the day {1,2,3} to represent {Morning, Afternoon, Evening}	DerivedFromOriginal
RestaurantID	Categorical	restaurant unique ID	Original
RestaurantLatitude	Numeric	latitude of the restaurant	Original
RestaurantLongitude	Numeric	longitude of the restaurant	Original
TrafficMorning	Ordinal	typical traffic in the morning {1,2,3} to represent {Green, Orange, Red}	DerivedFromOriginal
TrafficNoon	Ordinal	typical traffic in the noon {1,2,3} to represent {Green, Orange, Red}	DerivedFromOriginal
TrafficAfternoon	Ordinal	typical traffic in the afternoon {1,2,3} to represent {Green, Orange, Red}	DerivedFromOriginal
DistanceKm	Numeric	distance in meters/1000	DerivedFromOriginal
TimeMinutes	Numeric	time in minutes	Original
M1DeliveryCost	Numeric	cost of delivery in Columbian pesos (currency)	Original
M2ExpectedDeliveryTime	Numeric	expected delivery time in minutes	Original
M3MinChargeForOrdering	Numeric	minimum Charge Ordering in Columbian pesos (currency)	Original
M3LogMinChargeForOrdering	Numeric	natural logarithm of M3	DerivedFromOriginal
M4NumberOfComments	Numeric	number of customer comments	Original
M4LogNumberOfComments	Numeric	natural logarithm of M4	DerivedFromOriginal
M5DeliveryTimeFulfillment	Numeric	difference between M2 and TimeMinutes	DerivedFromOriginal
M5Class	Binary	class for M5; 0: Infeasible (M5<0) and 1: feasible (M5>=0)	DerivedFromOriginal
M6DeliveryCostPerKm	Numeric	M1DeliveryCost divided by DistanceKm	DerivedFromOriginal
M6Class	Binary	class for M6; 0: High (M5>2500) and 1: Acceptable (M6<=2500)	DerivedFromOriginal

To this end, a custom-tailored predictive data analytics workflow was developed in the study and formalized as a methodology.

Fig. 1 illustrates the steps of the developed and applied methodology. The first page of the workflow on the left, referred to as Page 1, details the steps of data preparation. The second page of the workflow on the right, referred to as Page 2, details the steps, which consist of Predictive Analytics and Visual Analytics.

A. Source Data

The methodology was developed specifically for the dataset acquired, relating to online delivery services.

As mentioned earlier, the dataset in this research was acquired from an earlier research study [4], whose data was published publicly and described as a separate paper [5-6]. The dataset was compiled to assess the effect of traffic conditions on the key performance indicators of 787 restaurants in Bogotá that offer food delivery services. The dataset contains the physical locations of restaurants and 4,296 customers in the city of Bogotá, Columbia, as well as the key performance indicators (KPIs) and traffic densities captured by Google Maps API during three time slices (morning, noon, afternoon) during Saturday.

The data was scraped by authors of the source studies using a custom-developed web scraping software, namely Agenty. The estimated travel times between the restaurant's physical location and the customers' addresses were also obtained through the Google Maps API. Since the website's information policy says that customer information must be kept private, the authors generated random samples of geographic points from Google Maps to fill for real customer addresses. Therefore, in our predictive analytics study, since they were simply generated, we did not use the location information for customers.

The original dataset contains 19,934 observations and 19 features in total, consisting of numeric, text, and categorical features. After the completion of data preparation, in our study, the features listed in Table 1 were eventually used in modeling and analysis. 93% of the observations in the original data were included in predictive analytics.

B. Data Preparation

Data preparation is the process of cleaning and transforming unprocessed data. It is the most important step before processing, and it often involves reformatting data, fixing data, and combining data sets to make the data more useful.

Data preparation is important for data professionals because it gets rid of any bias caused by bad data quality and ensures that any insights obtained are accurate and reliable. In this paper, our preparation includes handling irrelevant data, transforming categorical features, feature selection, and feature engineering.

1) Data Cleaning

The attribute number of comments is numeric, which means each restaurant receives a comment from the customer. However, there were irrelevant text values for some cells, and since values in those rows were possibly shifted, we deleted all those rows. We used filters in MS Excel to sort the data and get rid of the data that wasn't relevant. There were many other data cleaning steps applied, in accordance with a well-established taxonomy of dirty data [11].

2) Feature engineering

Feature engineering is a crucial step in machine learning data preparation. It uses existing data to generate new features (attributes, variables) that are not originally present in the dataset yet can be inferred from the data. Feature engineering can generate new features for both supervised and unsupervised learning, with the aim of simplifying and accelerating data transformations while improving model precision. To this end, transformation functions such as arithmetic, aggregation, and logic operators are commonly applied [11].

Besides systematically renaming the KPI readily available in the original data [5-6], new KPIs were also generated. Especially M5Class and M6Class, which did not exist in the original data, as they were analyzed using classification, yielded interesting insights.

Below are detailed description of some of the features corresponding to KPIs:

M5DeliveryTimeFulfillment: calculated as the difference between expected travel time quoted by the restaurant (in minutes) minus the travel time estimate provided by Google Maps API. This metric denotes of how much "overpromising" is done by the restaurant. If the value of M5DeliveryTimeFulfillment is negative, it is simply not possible for the restaurant to deliver on time, because the travel time TimeMinutes is already more than the quoted M2ExpectedDeliveryTime.

M5Class: The delivery time fulfillment class, which takes value 0 (Infeasible) if (M5<0) and 1(Feasible) if (M5>=0).

M6DeliveryCostPerKm: M1DeliveryCost divided by DistanceKm. This metric denotes how much "overpricing" is done by the restaurant. High values mean that the restaurant charges more per km. If a restaurant is overpricing much more than the majority of restaurants, it should be investigated, as high costs for the restaurant may also negatively affect the reputation of the OFD service.

M6Class: The DeliveryCostPerKm class, which takes value 0 (High) if (M5>2500) and 1 (Acceptable) if (M6<=2500). The threshold of 2500 approximately corresponds to the sample average plus one sample standard deviation.

3) Encoding Categorical Features

Encoding categorical data is the process of converting categorical data into an integer format so that the data has binary or ordinal values that can improve the models' predictive power [13].

Ordinal data is a type of data that is used to indicate the rank or order of certain items. This can be helpful when trying to determine the hierarchy of certain elements, or when trying to compare items that are not directly comparable.

As an example, three of the categorical attributes converted to ordinal integer values are TypicalTrafficMorning, TypicalTrafficNoon, and TypicalTrafficAfternoon. The conversions were $Green \rightarrow 1, Orange \rightarrow 2,$ and $Red \rightarrow 3.$ We used the IF function in Microsoft Excel to encode select categorical features in the original dataset as ordinal features in the prepared dataset.

Furthermore, Moment, which is used to take categorical text values, was also encoded using ordinal values of 1, 2, 3, to represent Morning, Noon, and Afternoon, respectively.

4) Feature Selection

Feature selection is the process of selecting a subset of features from a larger set of features. This can be done for a variety of reasons, such as reducing the dimensionality of the data, increasing the interpretability of the model, or improving the performance of the model.

As the curse of dimensionality reduction is a significant issue that may result in overfitting, feature selection is a crucial step in data preparation. Functional feature selection eliminates redundant and irrelevant features. As can be seen in Table 1, features that takes categorical values were omitted, as they were replaced with new features that take ordinal values, such as Moment. Furthermore, other exclusions were made, such as removing the feature travel time in seconds that came with the original data, as we already include TimeMinutes.

C. Visual Analytics

Visual analytics is a field of study that deals with the analysis of data using visual methods [14]. This includes the use of visual aids such as charts, graphs, and maps to help make sense of large data sets. The aim of visual analytics is to help people see and understand patterns and relationships in data that would be difficult to discern using traditional methods such as statistical analysis. Visual analytics is part of the methodology and was extensively applied in the study, even though not reported in the paper.

D. Predictive Analytics

Predictive analytics refers to the set of techniques applied for making predictions about unknown or future data instances, based on existing and historical data. Predictive analytics is enabled by increasing availability of large data sets, which are referred to as big data, as well as advances in machine learning (ML).

1) Classification

In machine learning, classification is a method of learning how to assign labels to data points. In classification, the target takes categorical values. This can be done using a variety of techniques but is typically done by training a classifier on a dataset of labeled data points and then using that classifier to predict the labels for new data points. There are many classification algorithms in machine learning, and 11 of these were employed in our study to predict two KPIs (M5Class and M6Class).

2) Regression

Regression analysis is a set of statistical techniques used to estimate relationships between a dependent variable where it has a numerical value and one or more independent variables. It can be used to predict the long-term relationship between variables and gauge how strongly a relationship exists between two variables [15].

In our study, we applied four regression algorithms to predict six KPIs that take numerical values.

3) Performance Comparison

In this paper, we consider AUC as the measure to evaluate the performance of the classification algorithms.

The AUC is a measure of how well a classifier can differentiate between two categories. It is also called the area under the curve, or AUC, for short. The AUC represents fifty percent the probability that a classifier will be able to correctly classify an instance of a positive class as positive, and an instance of a negative class as negative. A classifier with an AUC of 1.0 is considered perfect.

We utilize MSE and R square are among the measures we use to evaluate the performance of the applied regression algorithms. MSE (Mean Squared Error) estimates the mean of the squares of the errors, that is, the mean squared difference between the estimated values and the actual values, using sample data. R^2 (R-Square) represents the percentage of variability in a data set that is explained by a linear regression model.

For comparing the performances of different classification and regression algorithms, we present comparison tables. For comparing the performances of different classification algorithms, we present ROC curves, which are plots of the performance of a classifier system as its discrimination threshold is changed. The curve is created by plotting the true positive rate (TP rate) against the false positive rate (FP rate) at different threshold values. TP rate is also known as sensitivity, recall, or probability of detection in machine learning.

IV. Analysis and Results

The presented predictive analytics methodology extends earlier work with this type of data through predictive analytics. The methodology was applied thorougly to obtain a pleothera of results. However, due to space constraints, only sample representative analysis results will be included and explained. The results we describe will be mainly related to the analysis of the categorical class label M5Class, which is a KPI derived from the value of the numeric feature/KPI M5DeliveryTimeFulfilment. Since M5 is dependent on another KPI, M2ExpectedDeliveryTime, regression results for M2 are also presented as sample results. The analysis was conducted using the Orange Data Mining Toolbox [16].

A. Classification Analysis

Table 2 presents Classification performance comparison for M5Class. The applied classification algorithms are sorted with respect to decreasing AUC values. As the results of the classification techniques performance comparison table show, Random Forest had the best performance by area under the curve (AUC) = 0.97 and classification accuracy = 96%.

Table II. Classification performance comparison for M5Class.

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.97	0.96	0.95	0.95	0.96
Gradient Boosting	0.95	0.94	0.94	0.94	0.94
CN2 rule inducer	0.95	0.96	0.96	0.96	0.96
Neural Network	0.94	0.94	0.93	0.93	0.94
AdaBoost	0.94	0.97	0.97	0.97	0.97
Logistic Regression	0.92	0.93	0.93	0.93	0.93
Naive Bayes	0.89	0.88	0.89	0.91	0.88
kNN	0.85	0.93	0.92	0.92	0.93
Tree	0.84	0.95	0.95	0.95	0.95
SGD	0.72	0.92	0.92	0.92	0.92
SVM	0.66	0.74	0.78	0.84	0.74

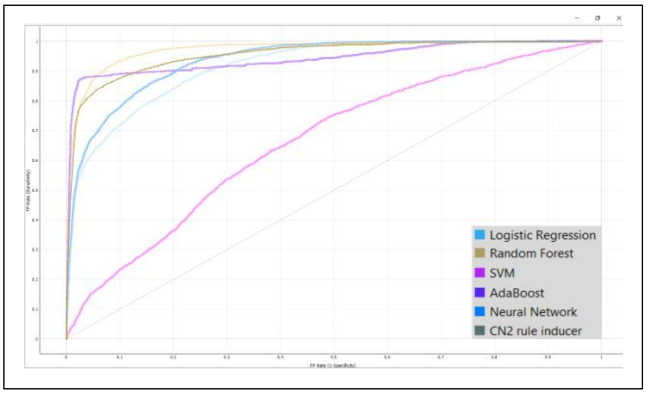

Fig. 2. ROC analysis for M5Class prediction.

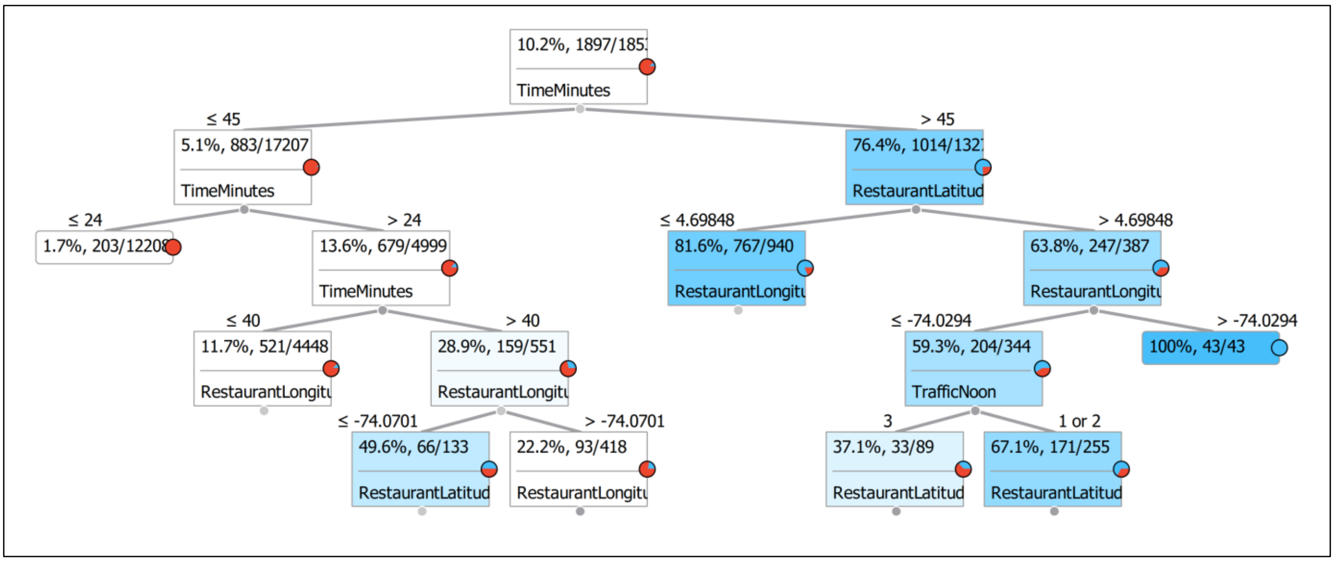

Fig. 3. Classification tree for M5Class.

The ROC curves in Figure 2 suggest that Random Forest (RF) (denoted with beige color) has the largest area under the curve in the ROC graph, followed by the CN2 rule inducer (denoted with dark green color).

Among the compared algorithms, the one that yielded a closed-form mathematical formula was logistic regression (LogRegr). For LogRegr, the equation for calculating the probability p that the delivery time fulfillment is feasible (M5Class =1) was derived as:

$$p = \frac{e^{-y}}{1+e^{-y}} \qquad (1)$$

where
y = 0.0449
− 0.0270×(TimeOfDay=1)
− 0.0682×(TimeOfDay=2)
+ 0.1375×(TimeOfDay=3)
+ 4.7096×(RestaurantLatitude)
+ 0.2090×(RestaurantLongitude)
− 0.0315×(TrafficMorning=1)
+ 0.0348×(TrafficMorning=2)
+ 0.0390×(TrafficMorning=3)
+ 0.2461×(TrafficNoon=1)
− 0.0505×(TrafficNoon=2)
− 0.1534×(TrafficNoon=3)
− 0.4012×(TrafficAfternoon=1)
− 0.1141×(TrafficAfternoon=2)
+ 0.5575×(TrafficAfternoon=3)
+ 0.0030×(DistanceKm)
− 0.1416×(TimeMinutes)

Figure 3 presents the classification tree for M5Class. Blue represents M5Class=0 and red represents M5Class=1.

In classification tree visualization, we particularly search for splits that result in significant changes in the proportion of observations for different classes. First such split is related to the value of Time Minutes being less than or equal to 45 or greater than 45, because the instances that satisfy TimeMinutes > 45 have visibly higher proportion of values for the class value M5Class=0. For the instances where TimeMinutes > 45, another such split rule is related to RestaurantLattitude, whose values larger than 4.698484 yield a visibly higher proportion of instances with class value M5Class=1. Yet another split finally results in the following rule:

```
IF
TimeinMinutes > 45
RestaurantLatitude > 4.69848
RestaurantLongitude > -74.0294
THEN
M5Class=0
```

This decision rule has 100% confidence, meaning that all the 43 instances in the dataset that satisfy the antecedent (IF ...) of the rule also satisfy the consequent (M5Class=0). The support of the rule, on the other hand, is quite low, being 43/18,534= 0.002320=0.2320%

B. *Regression*

Table 3 presents the comparison table for regression analysis of M2ExpectedDeliveryTime. The table suggests that Random Forest and Tree are the best machine learning algorithms with the lowest MSE, RMSE, MAE, and highest R^2 = 97 % values.

Table III. Regression performance comparison for the KPI M2ExpectedDeliveryTime. The applied regression algorithms are sorted with respect to decreasing MSE values.

Model	MSE	RMSE	MAE	R2
Random Forest	4.79	2.19	0.45	0.97
Tree	5.43	2.33	0.39	0.97
Neural Network	140.07	11.84	8.31	0.12
Linear Regression	155.52	12.47	8.41	0.02

The decision tree algorithm builds a tree for predicting the values of the target feature M2. Since the tree has many branches and leaves, and many rules correspond to the leaves, we provide a single rule as an example: One of the rules in the decision tree regression model, that is satisfied by 64 instances (observations) in the dataset, is the following:

```
IF
RestaurantLongitude <= -74.0688
RestaurantLongitude <= -74.0921
RestaurantLongitude <= -74.0923
RestaurantLongitude <= -74.1015
TrafficAfternoon = 2 or 3
RestaurantLongitude > -74.1768
RestaurantLongitude < -74.1688
RestaurantLongitude > -74.174
THEN
M2ExpectedDeliveryTime = 18.8 +- 4.7
```

This particular rule can be expressed succinctly as follows:

```
IF
RestaurantLongitude < -74.1688
RestaurantLongitude > -74.174
TrafficAfternoon = 2 or 3
THEN
M2ExpectedDeliveryTime = 18.8 +- 4.7
```

For the instances that satisfy the antecedent (IF …) of the rule, which depends mainly on the geographical longitude of the restaurant and the level of traffic in the afternoon (taking values of medium and high levels of traffic), the prediction is the mean value of 18.8, with a standard deviation of 4.7.

V. Conclusions and Future Work

In predictive analytics for OFD services, machine learning (ML) algorithms can be used to improve our understanding and prediction of KPIs. In this study, we applied RF and other ML algorithms for predicting six KPIs from OFD, as well as variants of these six KPIs. Random Forest (RF) was observed to consistently yield the best predictions in both classification (for estimating categorical attributes) and regression (for estimating numerical attributes). In the paper, as a theoretical contribution, the application of predictive data analytics for data from OFD services has been formally framed as a methodology, which can be adopted and extended to similar problems coming from OFD services. As a practical contribution, the methodology and its application yielded new types of insights that did not exist in the earlier related OFD studies.

References

[1] Li, C., Mirosa, M., & Bremer, P., "Review of online food delivery platforms and their impacts on sustainability," Sustainability, 12 (14), pp. 5528, 2020.

[2] Wei, C., Asian, S., Ertek, G., Hu, Z.-H., "Location-based pricing and channel selection in a supply chain: a case study from the food retail industry," Annals of Operations Research. https://doi.org/10.1007/s10479-018-3040-7, 2018.

[3] Alalwan, A. A. (2020). "Mobile food ordering apps: An empirical study of the factors affecting customer e-satisfaction and continued intention to reuse," International Journal of Information Management, 50, pp. 28-44, 2020.

[4] Correa, J. C., Garzón, W., Brooker, P., Sakarkar, G., Carranza, S. A., Yunado, L., & Rincón, A., "Evaluation of collaborative consumption of food delivery services through web mining techniques," Journal of Retailing and Consumer Services, 46, pp. 45-50, 2019.

[5] Correa, J.C. (2018). "Raw Data of A Web Mining Approach to Collaborative Consumption of Food Delivery Services," Mendeley Data, V1, doi: 10.17632/m9z9hw4nsc.1. Available under https://data.mendeley.com/datasets/m9z9hw4nsc/1. Accessed on Oct 22, 2022.

[6] Segura, M. A., & Correa, J. C.,"Data of collaborative consumption in online food delivery services," Data in Brief, 25, pp. 104007, 2019

[7] Teichert, T., Rezaei, S., & Correa, J. C., "Customers' experiences of fast food delivery services: uncovering the semantic core benefits, actual and augmented product by text mining," British Food Journal, 122(11), pp. 3513-3528, 2020.

[8] Modak, K. C., & Sinha, K., "A Logistic Regression Model of Customer Satisfaction for Online Food Delivery Services," 2019.

[9] Prasetyo, Y. T., Tanto, H., Mariyanto, M., Hanjaya, C., Young, M. N., Persada, et al., "Factors affecting customer satisfaction and loyalty in online food delivery service during the COVID-19 pandemic: Its relation with open innovation," Journal of Open Innovation: Technology, Market, and Complexity, 7(1), pp. 76, 2021.

[10] Cheng, C. C., Chang, Y. Y., & Chen, C. T., "Construction of a service quality scale for the online food delivery industry," International Journal of Hospitality Management, 95, pp. 102938, 2021.

[11] Kim, W., Choi, B. J., Hong, E. K., Kim, S. K., & Lee, D., A taxonomy of dirty data. "Data mining and knowledge discovery, " 7(1), pp. 81-99, 2003.

[12] Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., & Turaga, D. S., "Learning Feature Engineering for Classification," In IJCAI, pp. 2529-2535, August 2017 August.

[13] Dahouda, M. K., & Joe, I., "A deep-learned embedding technique for categorical features encoding," IEEE Access, 9, pp. 114381-114391, 2021.

[14] Ertek, G., Al-Kaabi, A., & Maghyereh, A.I., "Analytical Modeling and Empirical Analysis of Binary Options Strategies," Future Internet. 2022; 14(7):208. https://doi.org/10.3390/fi14070208, 2022. Available under https://www.mdpi.com/1999-5903/14/7/208. Accessed on October 22, 2022.

[15] Brook, R. J., & Arnold, G. C., Applied regression analysis and experimental design. CRC Pressi 2018.

[16] Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M. et al., "Orange: Data Mining Toolbox in Python," Journal of Machine Learning Research 14(Aug): pp. 2349−2353, 2013.