Applied Information Technology And Computer Science Vol. 4 No. 1 (2023) 1654-1667

AITCS

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/aitcs
e-ISSN :2773-5141

Explainable AI Using Data Robot to Identify Nutrition Values of Ordered Food

Nur Masiha Mohd Shariff[1], Mohammed Saeed Jawad[1*]

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/aitcs.2023.04.01.100
Received 14 June 2022; Accepted 07 June 2023; Available online 30 June 2023

Abstract: Explainable artificial intelligence (XAI) is an advanced form of machine learning that allows people to understand and trust machine learning algorithms' results. Users can drag and drop the dataset to test their model, which can be completed in about an hour. This paper aims to investigate and simulate the explainable AI using a Data robot to identify the nutrition value of ordered food by predicting the health condition features. The framework was created automatically in the data robot system and could select the target of your searching dataset to give the best model. This research will use a data robot as a platform. The reason the data robot was chosen is that it can use for making automation predictions and fitting graphs that show a certain detail in terms of an outlier, and visualizations that can easily interpret the result more understanding. Besides, data robots could assist you with a bunch of machine learning models that can be used for test prediction and deployment. The model is organized well in the data robot after the dataset has been implemented along with the target prediction. The performance of the metric is also recommended by the data robot that fits with the dataset itself. In this experiment, two types of models that work on binary classification have been selected named as Keras Slim Residual Neural Network Classifier using the Training Schedule Model and Generalized Additive2 Model. It would be comparing the performance accuracy at the end of the deployment of the dataset. It goes to show which model used has a better performance to explain the nutrition value of ordered food by predicting the health condition and could be implemented in the food industry in the future with more training and testing of different datasets categories and different machine learning models to verify the prediction performances.

Keywords: Artificial Intelligence, Data robot, Explainable AI, Nutrition value, Dataset food, McDonald's dataset, Data deployment, Binary Classification.

1. Introduction

Explainable artificial intelligence (XAI) is considered an AI system's interaction with humans to indicate the fundamental principles for an explainable AI system [1]. It also allows people to

Corresponding author: saeed@uthm.edu.my

understand and trust machine learning algorithms' results and output. The term "explainable AI" refers to a model's projected impact and probable biases. It aids in the evaluation of model accuracy, fairness, transparency, and outcomes in AI-assisted decision-making. When it comes to putting AI models into production, an organization's ability to explain AI is critical. AI machine learning also aids an organization's adoption of a responsible AI development strategy [2]. AI system should be able to explain its outputs and provide evidence to back up that explanation. Besides, AI systems have become one of the highest requirements to develop an automated system while providing logical information regarding certain things such as detecting a variety of food nutrition thus it can help any developers ensure that the system is working as expected and follow the standard requirement.

What is a data robot, then? A data robot is an automated machine learning system that uses data science methodologies to generate and deploy extremely accurate models. Data robots can help with data correctness by exploring millions of algorithms for deployment, data pre-processing stages, supplying features, and fine-tuning parameters for the optimal machine learning model. Users can also drag and drop the dataset to test their model, which can be completed in about an hour. The main objective is to investigate the features of the Data robot to identify the nutrition value of ordered food and determine which machine learning has better accuracy in analyzing the selected data. This objective gives huge different feedback and results for this project.

With the current advancements in each deep learning approach, it may be possible to create algorithms that generate greater accuracy and reduced computation using well-known data. There are some similar things and methods been used to classify the nutrition inside the food by using only machine learning and deep learning, not the data robot functions. However, the problem is that people are unconcerned about the nutrition value of their ordered food, which can have a positive or negative impact on their physical bodies, which may be deficient in certain nutrients, so it will be determined whether the food is healthy or not by looking at the nutrition value. This will be one of the worldwide knowledge areas for individuals who require nutrition in their daily lives and to support normal growth, hence lowering the risk of chronic disease and an unhealthy lifestyle. This research will focus on identifying nutrition values in each of portion food using a data robot so that users will get good accuracy and the proper diet could be done easily. So, it would conclude that this paper's aim is to investigate and simulate the explainable AI using a Data robot to identify the nutrition value of ordered food automatically.

This project research covers a wide range of topics from explainable AI to identifying the nutrition value of ordered food using data robots. To meet the project's objective, it will need to do much more research projects in terms of the explainable AI, how to train datasets in the data robot, and what output could be produced by using the data robot. This research involves how the process will run, what model is involved in the process for explainable AI, and the opportunities to explore the available model that existed in the data robot thus running an algorithm to identify the nutrition value of ordered food. The framework was created automatically in the data robot system and could select the target of your searching dataset to give the best model.

The data robot is essentially doing a quick raw data read and then going to perform exploratory data analysis. Different datasets would consist of different types of model deployment that are written in different programming languages. The data robot will be ready to make predictions based on what the target needs. In conclusion, this research aims to identify the nutrition values of ordered food by using a data robot. All the datasets chosen would be trained in the data robot of AI. This gives a lot of features model-like prediction features that would give more accurate analyses of the datasets and provide recommended models for deployment.

The framework was created automatically in the data robot system and could select the target of your searching dataset to give the best model. The target for this project is a health condition which is going to be a major part of knowing the nutrition value of ordered food. Based on the project structure, it will use a data robot to create a model to identify the nutrition value of ordered food by predicting through health conditions of McDonald's dataset. The dataset should have a minimum row requirement to be training that has a few types of variables. For this project, will use a dataset from McDonald's that has 25 features with 260 data points that are suited to be implemented in the model selected. The

framework of the model will use a data robot as a visual AI tool to represent the two different models. In addition, to create the model will use binary classification tasks. By filling the objective, it will create an environment for people to manage the process more efficiently and at the end of this project, people could get the result about the nutrition value of ordered food by predicting their health condition. To make it clearer to prove which model in the binary classification has the best performance, it will have one more model to make a difference in predicting the health condition that would result on the list of nutrition values of the ordered food.

This paper will provide an overview of the binary classification model such as Keras Slim Residual Neural Network Classifier using Training Schedule and Generalized additive2 Model to identify the nutrition value of ordered food through the prediction of health conditions. Section 2 reviews all the work related to the identified food nutrition while section 3 describes the methodology of the research that describes how the explainable AI can be optimized using a data robot to perform the task with the datasets and evaluation prediction.

2. Literature Review

2.1 Domain Case Study

Artificial intelligence (AI) has become one of the important and expert systems that would be used in engineering, science, education, finance, accounting, and much more field that existed over decades ago. Artificial intelligence (AI) is a broad word that refers to any type of computer software that performs human-like tasks such as learning, planning, and problem-solving [3]. AI has its methods which are the symbolic approach and the connectionist approach. The symbolic approach is known as a top-down approach that aims to reproduce intelligence by evaluating cognition in terms of symbol processing, regardless of the organic structure of the brain while the connectionist approach involves creating artificial neural networks and is known as the bottom-up approach.

Next, Explainable AI (XAI) is a system that could provide explanations to make its behavior more understandable to humans. Certain broad ideas may be used to assist the design of AI systems that are more successful and intelligible to humans: The XAI system should be able to explain its capabilities and understandings, as well as what it has done, what it is doing now, and what will happen next. It should also be able to reveal the critical information on which it is acting [4].

Data robot provides a model-agnostic framework that humans may use to evaluate outcomes for all specified models. Because the machine learning model offered takes appropriate tools and techniques, it may result in biased decision-making, which might provide additional information for documentation on how to describe the project to others. There are several features in the data robot that can instruct and investigate individuals to help them achieve their major purposes and decision results. Besides, a data robot is a leading end-to-end enterprise AI platform that automates and accelerates every step of the path from raw data to value [5]. Amazon Web Services (AWS) is used to host the cloud module, making it more adaptable and agile, as well as simplifying machine learning. With Data Robot, any application can deploy predictive models quickly and use them to increase value generation with only a few clicks. However, it also includes exportable prediction code, real-time prediction APIs, batch and native scoring, and exportable prediction code, among other ways for creating final predictive models.

To run the prediction of the selected target, which is a health condition, two different models had been chosen are Keras Slim Residual Neural Network Classifier using Training Schedule and Generalized Additive2 Model to make a comparison in terms of accuracy which model is better performance during the prediction. Neural networks are usually used for estimating some functions that consist big of data thus it has a numeric weight that can be tuned to run some learning. Neural networking is an optimizer to calculate the prediction and actual prediction. This model would be accepting any variable that is included in the dataset such as categorical, numeric, and text variables. One-hot encoding is needed to transform the categorical variable before sending it to the Keras for prediction. Besides, the numeric variables will undergo cleansing to make sure to disguise missing

value for data quality and convert all the numeric features into the same scale between -1 to 1 thus text variables are needed to convert raw text-filed into a document term matrix. Besides, all the variables in the dataset are required to be converted to tune the data before being transmitted to the Generalized Additive2 Model to make an understandable model based on Gradient Boosting Machine that approximates non-linear interactions before making a prediction.

2.2 Dataset

Kaggle is a platform for exchanging ideas, being inspired, competing against other data scientists, learning new knowledge, and coding methods, and witnessing real-world data science examples. Several datasets may be used for everything from video game sales to more sophisticated and essential data such as air pollution statistics. Because this data is actual and referenced, you may use it to train and test your models on real-world tasks. Data, code, community, inspiration, competition, and courses are just a few of Kaggle's many beneficial elements [6]. Users may use Kaggle to search and publish data sets, study, and construct models in a web-based data-science environment, collaborate with other data scientists and machine learning experts, and compete in data science contests. [7]. Several datasets can be used on Kaggle. This research used a dataset from McDonald's from Kaggle that provides a nutrition analysis of every menu item on the US McDonald's menu in delimited files (xlsx) format, including breakfast, beef & pork, chicken and fish, fries, salads, etc.

2.3 Related Works

There are many research studies on food nutrition prediction. We selected three related studies including predicting macronutrients of baby food using near-infrared spectroscopy and a deep learning approach [8], a machine learning-based approach to food recognition and nutrition estimation [9], and last but not least, research on a framework to estimate the nutritional value of food in real-time using deep learning technique [10]

Table 1: Comparison of Related Research Paper

Author(s), Year	Dataset	Technique	Advantage	Limitation
(Aulia et al., 2020)	Instant-porridge	• Using camera-based technique and sensor-based technique named SCiO • Used CNN and DBN model • Used k-fold cross validation to evaluate the best model	• Gives exact result on the macronutrient content in the baby food	• The changes of dropout would decrease the training performance
(Shen et al., 2020)	Food-101	• Use three different modules which are text data retrieval, text data training and training classification model using CNN • Word2Vec as tools for	• Simpler the ways to measure daily calories estimates to help people balancing their diet.	• Data sets and features are not sufficient and contain limited parameters.

		training text data		

Table 2: Comparison of Related Research Paper (cont.)

Author(s), Year	Dataset	Technique	Advantage	Limitation
(Yunus et al., 2019)	• Food-101 • Food-5k	• Using CNN model based on VGG-16, VGG-19, Inception-v3, Inception-v4, ResNet • Using Continuous Bag of Works (CBOW) & Skip-Gram Model (SGM) approach	• Provide mobile-based app to display the nutritional value of meal using image	• Not all applicable to use the features provided

3. Methodology/Framework

In this chapter, the study's findings will be extensively explored in this chapter. The first section of this chapter will cover-up in what research will use and how each of the phase-in research will conduct in this research project. Some research papers, journals and articles will be used to assist and steer this project's success. There are a few aspects that will be discussed later in the process.

3.1 Research Framework

This section will explain the overview of the whole process on how this research project will be a workout by following the designed research framework for the purpose to complete this research project. Figure shows the proposed model of the research for identify the nutrition food consist of four phases which are McDonald's dataset, pre-processing, prediction, evaluation, and the output.
The flowchart for this research refers figure. A suitable dataset needed to be implemented in the data robot to find classifications model. Dataset may be different formats and purposes then the dataset McDonald required for identifying nutrition values of ordered food by predicting either the health condition is healthy or not healthy. The next phases involve data preprocessing to make sure that the machine learning model could deploy the dataset more efficient. The data transformation and feature selection were being gripped. The dataset was applied with the binary classification model since it is practically involves predicting one of two more classes [11]. Finally, the training and testing dataset would be evaluated based on the different type of model.

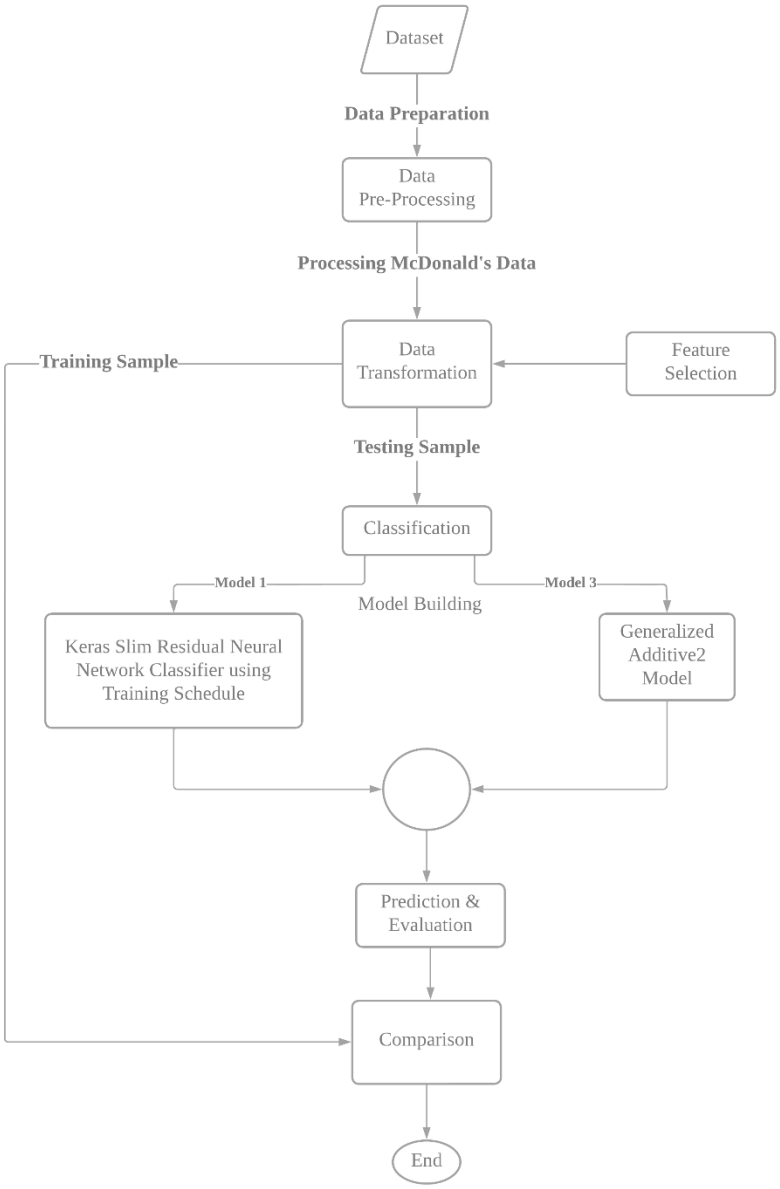

Figure 1: Flowchart of the entire research

3.1.1 Dataset

The classification was performed in the Data robot for prediction by using a raw dataset from the McDonald's that being extracted from Kaggle website. This dataset consists of 25 features which has 260 of datapoints that a few of variable types such as categorical, numeric and text. It also displayed the nutrition facts for each McDonalds's menu. (puneetk, 2017). This dataset allowed the data robot correctly to identify which classification model produced the highest accuracy of prediction. The dataset that contains nutrition values of McDonald is shown in Figure 2.

3.2.2 Data Pre-processing

Data preprocessing is required once the raw data has been implemented in the Data robot to perceive a suitable machine learning model by undergo data cleaning since the data generally contains missing value and it is a crucial step for creating a machine learning model which could perform a high level of accuracy.

Figure 2: Dataset that contains nutrition value of McDonald

Category	Item	Serving Size (g)	Calories	Calories from Fat	Total Fat	Total Fat (	Saturated	Saturated	Trans Fat	Cholester	Cholester	Sodium	Sodium (%	Carbohydi	Carbohydi	Dietary Fil	Dietary Fil	Sugars	Hea
Breakfast	Egg McMuffin	136	300	120	13	20	5	25	VERY LOW	260	87	750	31	31	10	4	17	3	Hea
Breakfast	Egg White Delight	135	250	70	8	12	3	15	VERY LOW	25	8	770	32	30	10	4	17	3	Hea
Breakfast	Sausage McMuffin	111	370	200	23	35	8	42	VERY LOW	45	15	780	33	29	10	4	17	2	Hea
Breakfast	Sausage McMuffin with Egg	161	450	250	28	43	10	52	VERY LOW	285	95	860	36	30	10	4	17	2	Hea
Breakfast	Sausage McMuffin with Egg Whites	161	400	210	23	35	8	42	VERY LOW	50	16	880	37	30	10	4	17	2	Hea
Breakfast	Steak & Egg McMuffin	185	430	210	23	36	9	46	LOW	300	100	960	40	31	10	4	18	3	Hea
Breakfast	Bacon, Egg & Cheese Biscuit (Regular Biscuit)	150	460	230	26	40	13	65	VERY LOW	250	83	1300	54	38	13	2	7	3	Hea
Breakfast	Bacon, Egg & Cheese Biscuit (Large Biscuit)	164	520	270	30	47	14	68	VERY LOW	250	83	1410	59	43	14	3	12	4	Hea
Breakfast	Bacon, Egg & Cheese Biscuit with Egg Whites (Regular Biscuit)	153	410	180	20	32	11	56	VERY LOW	35	11	1300	54	36	12	2	7	3	Hea
Breakfast	Bacon, Egg & Cheese Biscuit with Egg Whites (Large Biscuit)	167	470	220	25	38	12	59	VERY LOW	35	11	1420	59	42	14	3	12	4	Hea
Breakfast	Sausage Biscuit (Regular Biscuit)	117	430	240	27	42	12	62	VERY LOW	30	10	1080	45	34	11	2	6	2	Hea
Breakfast	Sausage Biscuit (Large Biscuit)	131	480	280	31	48	13	65	VERY LOW	30	10	1190	50	39	13	3	11	3	Hea
Breakfast	Sausage Biscuit with Egg (Regular Biscuit)	163	510	290	33	50	14	71	VERY LOW	250	83	1170	49	36	12	2	6	2	Hea
Breakfast	Sausage Biscuit with Egg (Large Biscuit)	177	570	330	37	57	15	74	VERY LOW	250	83	1280	53	42	14	3	11	3	Hea
Breakfast	Sausage Biscuit with Egg Whites (Regular Biscuit)	167	460	250	27	42	12	62	VERY LOW	35	11	1180	49	34	11	2	6	3	Hea
Breakfast	Sausage Biscuit with Egg Whites (Large Biscuit)	181	520	280	32	49	13	65	VERY LOW	35	11	1290	54	40	13	3	11	3	Hea
Breakfast	Southern Style Chicken Biscuit (Regular Biscuit)	143	410	180	20	31	8	41	VERY LOW	30	10	1180	49	41	14	2	6	3	Hea
Breakfast	Southern Style Chicken Biscuit (Large Biscuit)	157	470	220	24	37	9	45	VERY LOW	30	10	1290	54	46	15	3	11	4	Hea
Breakfast	Steak & Egg Biscuit (Regular Biscuit)	201	540	290	32	49	16	78	LOW	280	93	1470	61	38	13	2	8	3	Hea
Breakfast	Bacon, Egg & Cheese McGriddles	174	460	190	21	32	9	44	VERY LOW	250	84	1250	52	48	16	2	9	15	Hea
Breakfast	Bacon, Egg & Cheese McGriddles with Egg Whites	178	400	140	15	24	7	34	VERY LOW	35	11	1250	52	47	16	2	9	16	Hea
Breakfast	Sausage McGriddles	141	420	200	22	34	8	40	VERY LOW	35	11	1030	43	44	15	2	8	15	Hea
Breakfast	Sausage, Egg & Cheese McGriddles	201	550	280	31	48	12	61	VERY LOW	265	89	1320	55	48	16	2	9	15	Hea
Breakfast	Sausage, Egg & Cheese McGriddles with Egg Whites	205	500	230	26	40	10	52	VERY LOW	50	17	1320	55	46	15	2	9	15	Hea
Breakfast	Bacon, Egg & Cheese Bagel	197	620	280	31	48	11	56	LOW	275	92	1480	62	57	19	3	11	7	Hea
Breakfast	Bacon, Egg & Cheese Bagel with Egg Whites	201	570	230	25	39	9	45	LOW	60	20	1480	62	55	18	3	12	8	Hea
Breakfast	Steak, Egg & Cheese Bagel	241	670	310	35	53	13	63	MEDIUM	295	99	1510	63	56	19	3	12	7	Hea
Breakfast	Big Breakfast (Regular Biscuit)	269	740	430	48	73	17	87	VERY LOW	555	185	1560	65	51	17	3	12	3	Hea
Breakfast	Big Breakfast (Large Biscuit)	283	800	470	52	80	18	90	VERY LOW	555	185	1680	70	56	19	4	17	3	Hea
Breakfast	Big Breakfast with Egg Whites (Regular Biscuit)	272	640	330	37	57	14	69	VERY LOW	35	12	1590	66	50	17	3	12	3	Hea
Breakfast	Big Breakfast with Egg Whites (Large Biscuit)	286	690	370	41	63	14	72	VERY LOW	35	12	1700	71	55	18	4	17	4	Hea
Breakfast	Big Breakfast with Hotcakes (Regular Biscuit)	420	1090	510	56	87	19	96	VERY LOW	575	192	2150	90	111	37	6	23	17	Hea
Breakfast	Big Breakfast with Hotcakes (Large Biscuit)	434	1150	540	60	93	20	100	VERY LOW	575	192	2260	94	116	39	7	28	17	Hea
Breakfast	Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit)	423	990	410	46	70	16	78	VERY LOW	55	19	2170	91	110	37	6	23	17	Hea
Breakfast	Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)	437	1050	450	50	77	16	81	VERY LOW	55	19	2290	95	115	38	7	28	18	Hea
Breakfast	Hotcakes	151	350	80	9	13	2	9	VERY LOW	20	7	590	24	60	20	3	10	14	Hea
Breakfast	Hotcakes and Sausage	192	520	210	24	37	7	36	VERY LOW	50	17	930	39	61	20	3	10	14	Hea
Breakfast	Sausage Burrito	111	300	150	16	25	7	33	VERY LOW	115	38	790	33	26	9	1	5	2	Hea

3.2.3 Feature Selection

Adding features to the dataset can help the machine learning model be more precise by limiting on certain aspects that don't contribute anything. For example, if trying to predict health condition of people who ate the McDonald, serving size in gram may be important, but sometime the existing of vitamin C might be not too applicable. Besides, data robot would be able to show feature association such as association between the sugars and health condition which is if sugars intake is high, so the health conditions also high but it does not have stronger association.

3.2.4 Model Building

In this research project, there are two machine learning model has been selected to extract insights from the dataset to make better decisions. Also, it could provide some information of the model such as the model blueprints that consists relevant insight to ensure that the model is accurate to be used for the McDonald's dataset. This research used data robot as main software to build, train and deploy the best machine learning model for McDonald's dataset after the target has been selected which is health condition. This is the way of data robot execute the model without intimidating mathematics and without any coding as well.

For modelling processing, a target needed in this research is called health_condition as indicator to the data robot to specify the appropriate metric. Since this research is using binary classification, the formula shown in Eq.1 where p is the predicted value of y.

$$-(ylog(p) + (1-y)\log(1-p)) \qquad Eq.1$$

Area Under Curve (AUC) is being implemented in this McDonald's dataset for scoring models which is usually automatically chosen in data robot. It is to measure the inaccuracy of predicted probabilities. Data robot also provide insight to analyze the visualization on tree-based variable important that show a ranking from the most important variable until the lowest one in the model that could help on determine either the nutrition value in the food is healthy or not healthy.

3.2.5 Prediction & Evaluation Phase

In this research project, the main measure for the two types of machine learning model is the accuracy of performance. The performance of the binary classification is evaluated based on the following parameters:

- True Positive (TP): The health condition is unhealthy, so the model predicts unhealthy
- False Positive (FP): The health condition is healthy, but the model predicts unhealthy
- True Negative (TN): The health condition is healthy, so the model predicts as healthy
- False Negative (FN): The health condition is unhealthy, so the model predicts as healthy

Accuracy for the matrix can be calculated in the confusion matrix on these values' parameter using this following formula Eq. 2:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad Eq. 2$$

Next, Area Under Curve (AUC) is widely used in binary classification problem for evaluation that can be define in the Eq. 3, Eq. 4, Eq. 5 below

$$True\ Positive\ Rate = \frac{True\ Positive}{False\ Negative + True\ Positive} \qquad Eq. 3$$

True Positive Rate (Sensitivity): It would consider as positive data points

$$True\ Negative\ Rate = \frac{True\ Positive}{True\ Negative + False\ Positive} \qquad Eq. 4$$

True Negative Rate (Specificity): It would consider as negative data points

$$False\ Positive\ Rate = \frac{False\ Positive}{True\ Negative + False\ Positive} \qquad Eq. 5$$

Where True Positive Rate (Sensitivity) would consider as positive data points while True Negative Rate (Specificity) would be considered as negative data points and False Positive Rate would be considering as mistakenly positive, but all data points are negative.

Besides, this data robot also could provide information regarding F1 Score, Precision and Recall measurement that could be calculated in the formula shown in Equation. It can be concluded that when the high precision but lower recall, it would give the most accurate and the much greater F1 Score, the better performance of the model. [12]

3.3 Hardware and Software specification

To ensure the experiment run smoothly without any problem, hardware is the most important to be used to achieve the expected results. The hardware used to run this experiment is stated in table while or software requirement, Data robot website as a tool to implement the dataset.

Table 3: Hardware Requirement

Hardware	Descriptions	
	Windows Edition	System
MateBook D 15	Window 10 Home Single Language	Processor: AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz
		Installed RAM: 8.00 GB
		System Type: 64-bit operating system, x64-based processor

4. Results and Discussion

The results and discussion section presents data and analysis of the study. This section can be organized based on the stated objectives, the chronological timeline, different case groupings,

different experimental configurations, or any logical order as deemed appropriate. The interface result, it allows us to know the details about each of the features in the data robot so that the data robot can function successfully as planned. In this part, we only do the implementation for the insert the McDonald's nutrition dataset, play along with the features inside the data robot, and ensure all the features could be displayed to conclude the information regarding the dataset thus ensuring the best model that going to give the best performance. The result of this research highlighted the accuracy of health condition to identify nutrition values of the ordered food.

4.1 Proposed Solution

Some of the categorization techniques that will be used in this research project are binary classifications. Data robot gives an automatically the best and accurate model on the top, the least accurate at the bottom thus the best model suited with the dataset will be selected to be deploy. Data robot has flagged the Keras Slim Neural Network as recommended for deployment because it has accuracy almost equal to the most accurate model and runs faster with prediction time is 0.76 second. The table below shows the comparison between the two-model used for this experiment. Each of the model supposed to give an accuracy result as a solution for explainable AI.

Table 4: Comparison between the two model for experiment

Model	Accuracy	F1 Score	Positive Predictive Value (Precision)	Prediction Time (s)
Keras Slim Residual Neural Network Classifier using Training Schedule	1	1	1	0.76
Generalized Additive2 Model	0.9744	0.973	0.9474	0.72

4.3 Experimental Setup

Since this experiment is mainly used data robot as a platform to generate and analyze data accuracy prediction, that was the minimum requirement before starting this experiment. This is no coding environment since all the performance has been made behind the scenes of data robot itself. Besides, once the dataset being uploaded, data robot needed to perform the machine learning classification model to proceed with the test prediction and deployment phases. With all the dataset that following the specified in the data robot in term of row and column, this research can start making experiment to get the results.

4.4 Result and Discussion

After do some analysis in this part, a data has been discovered the results of prediction depicted in the table. Table shown a summarize of the characteristic of the datasets used in the prediction experiment.

Table 5: Characteristic of dataset

Dataset	No. of Features	Datapoints	Variable Type	
McDonald	25	260	Categorical	5
			Numeric	19
			Text	1

4.4.1 Test Prediction Using Two Model

In the classification process, the results from the features extraction are transformed into feature vector in a comma-separated value (CSV) file which consists of the McDonald datasets respectively. The vector contains values 0 or 1, which points its presence or absence in the sample. There are data in decimal value as well but already being transformed as single value. For this work, the vector will end will classify the sample as either 0 as healthy conditions or 1 as an unhealthy condition to identify the list of nutrition values in each ordered food. Then we used the two different model to clarify the dataset based on the given features. With the good accuracy, it can be concluded that which model could be used in explainable AI in food industry.

Next, table 5 and table 6 shown an example test prediction from the dataset that has a binary classification problem where data robot estimates the probability of unhealthy condition as 1 since the positive class already being decided is Unhealthy while predicted label 0 as healthy condition. This prediction result using two different model can be made after setting up the prediction threshold as 0.5. any prediction scores that exceed from this threshold will be assigned to the positive class which is unhealthy.

Table 6: Example prediction result of probability in the positive class and negative class of Generalized Additive2 Model

Row_id	Item	Health_condition	Trans Fat	Saturated	Prediction	Predicted Label
105	Oatmeal Raisin Cookie	Healthy	VERY LOW	3	5.95E-05	0
106	Kids Ice Cream Cone	Healthy	VERY LOW	1	0.000111	0
107	Hot Fudge Sundae	Unhealthy	VERY LOW	7	0.999988	1
108	Hot Caramel Sundae	Unhealthy	VERY LOW	5	1	1

Table 7: Example prediction result of probability in the positive class and negative class of Keras Slim Neural Network Model

Row_id	Item	Health_condition	Trans Fat	Saturated	Prediction	Predicted Label
105	Oatmeal Raisin Cookie	Healthy	VERY LOW	3	0.003532037	0
106	Kids Ice Cream Cone	Healthy	VERY LOW	1	0.010855364	0
107	Hot Fudge Sundae	Unhealthy	VERY LOW	7	0.991368353	1
108	Hot Caramel Sundae	Unhealthy	VERY LOW	5	0.995348752	1

4.4.2 Nutrition Value Prediction Result Analysis

In the experiment, the performance metric of accuracy value, and Confusion matrix are evaluated for McDonald's dataset on Keras Slim model and Generalized Addictive2 Model. To measure the effectiveness of the nutrition value prediction using binary classification, we consider using these performance metrics. After do some exploration, the accuracy of Keras Slim is 100 percent and 97.4 percent for Generalized Addictive2 Model. This shows that Keras Slim provides an accurate and reliable accuracy mainly for various types of data in classification. If the accuracy value is high, the performance of the proposed model is very effective to identify the nutrition value by targeting the health conditions either healthy or not healthy.

Besides, an accuracy has been discovered and results of the analysis depicted by the table below. The table 7 show the prediction information that been used to identify the nutrition values in ordered food by predicting the health condition. The analysis implement two type of machine learning models and one types of data set which is McDonald. This prediction would be tested in each model to see the different variation calculation. Therefore, this experiment will focus on the accuracy of the model after being deploy using AI app template that undergo the predictions with the data robot. This table shows the result of prediction after being deploy using predictor application in the data robot. This experiment targeted the health condition as the main prediction so that it could provide the information on nutrition values of ordered food.

Table 8: Prediction for breakfast that has actual healthy conditions

No.	Prediction Information	Model Prediction	
		Keras Slim Neural Network Model	Generalized Additive2 Model
1	Category: Breakfast Item: Egg McMuffin Serving Size (g): 136 Carbohydrates: 31 Calories: 300	Actual: 1.56%	Actual: 0.00%
	Category: Breakfast Item: Egg McMuffin Serving Size (g): 500 Carbohydrates: 56 Calories: 798	Prediction: 2.93%	Prediction: 0.00%

However, in the early training, a threshold has been set as 0.5 where if the value of prediction is more than 0.5, it could be grouped in the unhealthy group since the positive class has been made which is an unhealthy condition. Based on the table above, the actual conditions are healthy. After doing the prediction, both the models interpret the overall result as healthy. This concludes that this prediction could be added as a True Negative (TN) because the actual health condition is healthy, so the model predicts a healthy condition. These results show in table 7 that the Egg McMuffin could be eaten because it is still in the group of healthy foods.

Next, the table 8 shows predictions for breakfast on the Fruit & Maple Oatmeal which actual result of health condition is unhealthy for people. This results in the True Positive (TP) where the health condition is still unhealthy that produced by the model prediction also unhealthy condition. So, whoever going to eat Fruit & Maple Oatmeal dessert for daily consumption, could result in unhealthy conditions for a long period. This becomes unhealthy because the value of the prediction exceeds 50%, and it can be added to the positive class gradually.

Table 9: Prediction for breakfast that has actual unhealthy conditions

No.	Prediction Information	Model Prediction	
		Keras Slim Neural Network Model	Generalized Additive2 Model
2	Category: Breakfast Item: Fruit & Maple Oatmeal Serving Size (g): 251 Carbohydrates: 58 Calories: 290	Actual: 92.03%	Actual: 99.92%
	Category: Breakfast Item: Fruit & Maple Oatmeal Serving Size (g): 300 Carbohydrates: 67 Calories: 126	Prediction: 78.31%	Prediction: 99.90%

Table 10: Prediction for desserts that has actual unhealthy conditions

No.	Prediction Information	Model Prediction	
		Keras Slim Neural Network Model	Generalized Additive2 Model
4	Category: Desserts Item: Hot Caramel Sundae Serving Size (g): 182 Carbohydrate: 60 Calories: 340	Actual: 99.53%	Actual: 100%
	Category: Desserts Item: Hot Caramel Sundae Serving Size (g): 14 Carbohydrate: 64 Calories: 693	Prediction: 99.95%	Prediction: 100%

The table 9 shows precisely that both models get unhealthy conditions that result in the True Positive (TP) where when the actual data is an unhealthy condition, the model prediction shows the unhealthy condition range on between 90% and 100% of accuracy. In a conclusion, the Generalized Additive2 Model has a better performance than Keras Slim Neural Network in predicting the Hot Caramel Sundae condition as unhealthy condition thus people would know their ordered food by seeing the accuracy of the dataset.

5. Conclusion

This research project mainly used Artificial Intelligence which is quite advance for human to implement as their works in the industry. Moreover, this project was being developed to identify the nutrition value of ordered food by predicting the targeted features named health condition so that it could give a proper information once the prediction has done. All of this would achieve a better performance with the help of listed model that been trained suitable according to the McDonald's dataset. The capabilities of the data robot were examined based on the data obtained from a certain period. The results should be indicated from the data robot and related dataset to produce a suitable model for analyzing the accuracy and meaning of the prediction and shows the top performance model for the dataset itself. All the mean must be predicted from the prediction experiment from the dataset in terms of nutrition values that consist of details of calories, fat etc. The main challenge in this study is the size of the dataset and exploring a suitable algorithm for the numerical dataset. The findings of this study are hoped to be useful in helping the government and authorities to make an early strategy

to handle any disastrous events. Some of the datasets would not be uploading if there are not enough rows and not all the datasets could produce ROC curves. There are some limitations that was identified in this research which is the dataset has many classes and items that needs to identify thus it makes it hard to perform the prediction because not following the rues of dataset requirement to uploading to the data robot. The trial package also being given but only could be used for 16 days which make the work kindly stopped for a moment to do the exploration in predicting the dataset.

In the future, this research should be continued with more training and testing of different datasets categories and different machine learning model to verify the prediction performance. Besides, this is no code segment which valuable to be used in any industry especially in the McDonald branches. However, to make all the model more information and could provide the prediction explanation, at least it could state how the dataset should be added to compute the data robot to get a better result in confusion matrix and accuracy of the targeted feature.

Acknowledgment

I would like to express my special thanks to my supervisor Dr. Mohammed Saeed Jawad as well as to my parents, siblings, and friends for their support.

References

[1] P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, and M. A. Przybocki, "Four Principles of Explainable Artificial Intelligence," Aug. 2020, doi: 10.6028/nist.ir.8312-draft.

[2] "Explainable AI," *Ibm.com*, 2021. https://www.ibm.com/watson/explainable-ai (accessed Dec. 30, 2021).

[3] A. Uzialko, "How Artificial Intelligence Will Transform Business," *Business News Daily*, Apr. 22, 2019. https://www.businessnewsdaily.com/9402-artificial-intelligence-business-trends.html (accessed Dec. 30, 2021).

[4] P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, and M. A. Przybocki, "Four Principles of Explainable Artificial Intelligence," Aug. 2020, doi: 10.6028/nist.ir.8312-draft.

[5] "DataRobot," *DataRobot*, Jun. 25, 2021. https://www.datarobot.com/sg/ (accessed Dec. 30, 2021).

[6] M. Przybyla, "Why Does Everyone Use Kaggle? - Towards Data Science," *Medium*, Jul. 09, 2020. https://towardsdatascience.com/why-does-everyone-use-kaggle-db1bdf1f1b1a (accessed Dec. 30, 2021).

[7] Wikipedia Contributors, "Kaggle," *Wikipedia*, Nov. 06, 2021. https://en.wikipedia.org/wiki/Kaggle (accessed Dec. 30, 2021).

[8] Aulia, M. N., Khodra, M. L., & Koesoema, A. P. (2020). Predicting Macronutrient of Baby Food using Near-infrared Spectroscopy and Deep Learning Approach. *IOP Conference Series: Materials Science and Engineering*, *803*(1). https://doi.org/10.1088/1757-899X/803/1/012019

[9] R. Yunus *et al.*, "A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques," *IEEE Access*, vol. 7, pp. 2643–2652, 2019, doi: 10.1109/access.2018.2879117.

[10] Z. Shen, A. Shehzad, S. Chen, H. Sun, and J. Liu, "Machine Learning Based Approach on Food Recognition and Nutrition Estimation," *Procedia Computer Science*, vol. 174, pp. 448–453, 2020, doi: 10.1016/j.procs.2020.06.113.

[11] Brownlee, J. (2020). 4 Types of Classification Tasks in Machine Learning. *Machine Learning Mastery*, 1. https://machinelearningmastery.com/types-of-classification-in-machine-learning/

[12] Aditya, M. (2018). Metrics to Evaluate your Machine Learning Algorithm | by Aditya Mishra | Towards Data Science. *Towards Data Science*, 1–8. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234