APPLICATION OF RANDOM FOREST ALGORITHM ON WATCH PRICE PREDICTION SYSTEM USING FRAMEWORK FLASK

Dzakiyyatul Kirom Dalimunthe[1], Raden Bagus Fajriya Hakim[2*]

[1,2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia
Jl. Kaliurang Km 14.5, Yogyakarta, 55584, Indonesia

Corresponding author's e-mail:* 986110101@uii.ac.id

ABSTRACT

In the modern era like today, watches not only function as timepieces but have become a fashion trend for the community, especially teenagers. The increasing market demand for watches opens up opportunities for counterfeit watch sellers to sell their products by claiming that the watches they sell are genuine watches by offering relatively lower prices compared to genuine watches. This is very detrimental to consumers and also the watch industry. To minimize fraud committed by fake watch sellers, it is necessary to know the price of the original watch in advance before buying the desired watch. Therefore, the purpose of this study is to predict the price of watches using the Random Forest method and will be developed into a web system using the Framework Flask. The results of the study using 3337 trees obtained an accuracy rate of 84,98% with a MAPE of 15,02%. The most influential variable on the price of watches is the material variable, with the level of importance obtained at 0,359. After getting the best model, the model is then developed into a web system using the help of the Framework Flask and Heroku, which can later be accessed online.

1. INTRODUCTION

Watches are a tool that works as a timepiece so that people are always on time in doing everything [1]. However, in the modern era like today, watches do not only function as a tool to show time but have become a fashion trend for society, especially teenagers. People in this modern era are very concerned about the appearance and fashion used, not to mention complementary accessories to support the appearance, that is why watches are here to give a touch in perfecting their appearance and are also very suitable for use in daily activities, both formal and informal. People also believe that using branded watches can improve the quality of appearance to be more elegant, show one's personality, and show one's social class in the eyes of others. This is what spurs watch entrepreneurs to continue to innovate by presenting watches with the latest models that have uses that are tailored to watch users but still pay attention to the elegant elements of the watch [2].

According to Jhonny Thio Doran as the Board of Director of the Doran Group, watch sales have experienced a sharp spike during the Covid-19 pandemic, especially in sports watches. This is because the pattern of society is starting to change by leading a healthy life, such as diligently exercising, starting from running, cycling, and various other types of sports. The increase in sales of watches, one of which is experienced by Garmin brand watches. This watch from the United States has increased during the Covid-19 pandemic to reach 50%; it started in early April and May and continued to increase from 30% to 50%, especially at Garmin, which caused frequent out of stock, due to high orders during the Covid-19 pandemic [3].

The high market demand for watches makes the number of sellers of fake watches increase. According to Customs, the amount of counterfeiting and piracy worldwide reaches US\$ 200 – 360 billion per year or 5% – 7% of international trade. They market their products by claiming that the watches they sell are genuine high-quality watches by providing relatively lower prices than the original watches. If consumers can't tell the difference between a genuine watch and a fake one, they will be tempted by the seller's offer of a fake watch that offers quality at a relatively lower price [4]. However, when the watch is used, the skin becomes irritated, such as rashes and itching after a few weeks of use. Then when the watch is exposed to water, the machine is immediately damaged, and the watch can no longer be used [5]. That is what makes consumers lose confidence in watch products with certain brands, even though they do not know that the watches they have are fake watches. In addition to harming consumers, the watch industry that produces these brands will also experience huge losses, where it will have an impact on reducing the workforce, which will increase the unemployment rate [4].

To minimize fraud by sellers of fake watches, it is necessary to know the price of the desired original watch before going to the store in question or buying online. Therefore, in this study, a prototype watch prediction system will be made using the Random Forest method and Framework Flask.

2. RESEARCH METHODS

2.1. Logistic Regression

Logistic regression is a regression model used to see the effect of independent variables in the form of continuous and categorical data on the dependent variable in the form of categorical data. Logistic regression consists of binary, multinomial, and ordinal logistic regression. In binary logistic regression, the dependent variable consists of two categories, namely 0 and 1 [6]. If the dependent variable has two categories with values 0 and 1, a binary logistic regression model is used in its calculations, where this model follows a Bernoulli distribution, as follows [7].

Definition 1. π_i *is the probability of occurrence to-i,* y_i *is random variable to-i which consists of 0 and 1* [10].

$$f(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \qquad (1)$$

2.2. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a method found in the supervised algorithm, where this algorithm aims to obtain new patterns from data. KNN regression is an algorithm that provides an introduction to the K-nearest regression used to predict the output value [8].

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a very popular artificial intelligence-based method, and this method has advantages in optimizing pattern recognition systems with good generalization capabilities. SVM can also be used to predict data. This method can also be used in the case of non-linear data with a large number of input data and regression, where the Support Vector Regression (SVR) method contained in the SVM method will be used [9]. SVR itself is a development of the SVM method in the case of regression. SVR has a goal to find a function f(x) as a hyperplane (separation line) in the form of a regression function that corresponds to all input data with an error that is made as small as possible [10].

2.4. Decision Tree

A Decision Tree is a tree-like flowchart where each node represents a test on an attribute, each branch represents the results of the test, and the leaf nodes represent the classes. Decision Tree functions to find hidden models of data with a target variable and explore data that has passed the preprocessing stage so that it can be used to divide large data sets into smaller data sets by taking into account the destination variable [11].

A Decision Tree is a data structure consisting of nodes and edges. Decision Trees have been used in various fields of science, such as the health sector, which uses decision trees to diagnose patient diseases, psychology for theory in decision making, computer science in data structures, and other fields of science [12].

2.5. CART (Classification and Regression Tree)

CART (Classification and Regression Tree) is an algorithm of data exploration techniques, namely the Decision Tree technique. CART produces a regression tree for the dependent variable with continuous data and a classification tree for the dependent variable with categorical data. The main goal of CART is to obtain a model that is as accurate as possible for classifying [13]. This method is included in the non-parametric method, so it does not require assumptions in the analysis process.

2.6. Random Forest

Random Forest is a method developed from the CART (Classification and Regression Tree) method by applying the bootstrap aggregating (bagging) and random feature selection methods, which were first introduced by Breiman. This method has many trees formed so as to form a forest (forest), then the analysis is carried out on the number of trees that are formed [14]. Random Forest has three methods, namely General Random Forest, Conditional Random Forest, and Improve Results of Logistics Using Random Forest. When compared to the bagging method and the boosting method, random forest is included in the new classification technique [15].

The CART method is used in the calculation process in building a decision tree, where this method uses information gain in measuring an attribute selection that will be used at each node in a tree (tree) [16]. Where is the formula used in obtaining the information gain value, that is:

Definition 2. *n is the number of target class, p_i is Proportion of class i to partition D* [16].
$$Gain(A) = Info(D) - Info_A(D) \tag{2}$$

The value of $Info(D)$ is obtained by using the following formula:
$$Info(D) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3}$$

Then the value of $Info_A(D)$ can be obtained by using a formula like the following:

Definition 3. *v is Number of partitions, D_j is Total partition to- j, D is total rows on all partitions* [16].

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{4}$$

2.7. Data Source and Data Collection Technique

The data used in this study is secondary data taken using the web scraping method, where the data taken is all watch sales data contained on the "machtwatch" website page.

The data in this study were taken using the help of Data Miner by taking data in the form of a web scraping method. Web Scraping is a technique to collect information from a website automatically without having to copy the information manually. The purpose of this method is to find certain information and collect it into a new web [17].

2.8. Research Variable

Table 1. Definition of Research Variables

Variable	Variable Definition	Scale
Price	The selling price of the watch is based on the specifications of each watch	Numerical
Brand	The mark worn by the watchmaker for the watches produced	Categorical
Series	The type or types used to distinguish between watch brand models	Categorical
Model No.	The type of number used to distinguish the series on the watch brand	Categorical
Material	Materials used in the manufacture of watches	Categorical
Gender	To differentiate watches by gender	Categorical

2.9. Analysis Method

There are several methods used in the watch data analysis process, that is:

1. Scraping data that aims to collect watch sales data contained on the "machtwatch" website using Data Miner.
2. Preprocessing data that aims to clean up data that contains noise and contains empty data in order to produce data with variables that are ready to be used.
3. Descriptive statistical analysis used to find out the general description of the variables in the watch data using Microsoft Excel 2016 software.
4. Random Forest analysis used to find out the best modeling in watch price predictions, as well as how accurately the model is formed, is done with the help of Jupyter Notebook software found on Anaconda.
5. Predictive models formed in machine learning will be deployed into a system using the framework flask, which aims to connect the model with localhost with Visual Studio Code and Command Prompt tools.
6. The prediction model is developed into a web that can be accessed online using Git, Command Prompt, Visual Studio Code, and Heroku.

3. RESULTS AND DISCUSSION

3.1. Descriptive Analysis

Descriptive analysis is needed to provide an overview of the variables studied, namely the dependent variable in the form of price and the independent variable in the form of brand, series, model no., material, and gender.

The following is the percentage of watch product availability by brand on the machtwatch website. It can be seen that the most widely sold products on the website are products, with the Casio brand with a product availability percentage of 36%, then in the second position, there are products with the Alexandre Christie and Seiko brands, with each product availability percentage of 20%, followed by Fossil with 7%

product availability, Tissot and Alba 4%, Orient and Expedition 3%, Aigner 2%, and the last one is Citizen brand with 1%. More clearly, see the following picture.
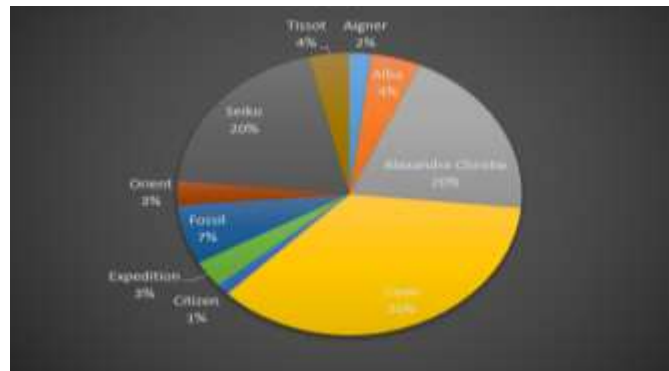

Figure 1. **Percentage of Brands Availability**

Then we can see the number of available types of watches based on movement, style, and needs on the machtwatch website. It can be seen in the following image that the most widely available types of watches on the website are watches with Quartz movement with a total availability of 1357 products, then the second type is found in Standard watches with an availability of 1161 products, in the third position there are types of watches. Classic watches with the availability of 1102 products, then Chronograph as many as 620 products, digital as many as 588 products, Automatic as many as 555 products, Sportwatch as many as 537 products, Solar as many as 127 products, Eco Drive as many as 66 products, Smartwatch and Diver as many as six products, then the last is a type of Pilot watches as many as three products. More clearly, see the following picture.
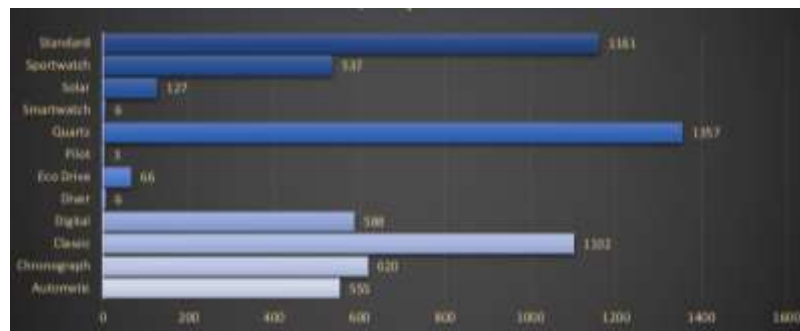

Figure 2. **Total Availability of Watch Type**

Each watch product has a different price depending on the watch brand. Therefore, we will look at the average watch price based on each brand. It can be seen in the following picture that the highest average watch price is found in watches with the Tissot brand of Rp. 9.698.712, then the second is found in products with the Aigner brand of Rp. 8.997.789, then products with the Citizen brand of Rp. 5.150.500, with the Orient brand of Rp. 4.522.653, the product with the Seiko brand of Rp. 4.499.665, the product with the Fossil brand of Rp. 1.747.774, the product with the Expedition brand of Rp. 1.660.077, the product with the Casio brand of Rp. 1.280.783, the product with the Alexandre Christie brand of Rp. 878.390, and The lowest average watch price is found in products with the Alba brand of IDR 724.847. More clearly, see the following picture.
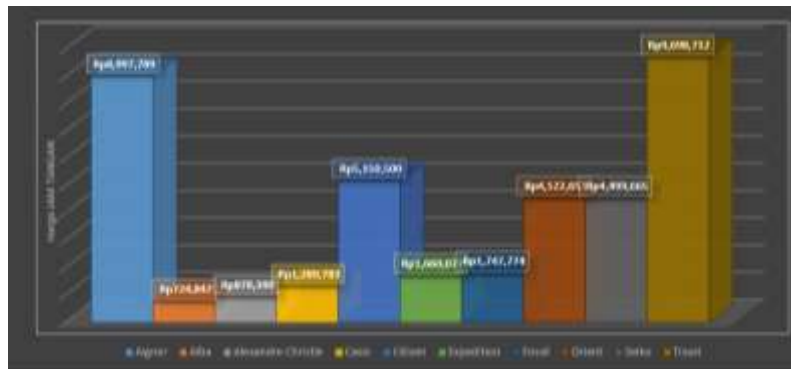
Figure 3. Average Watch Price by Brand

3.2.   Encoding Labels

Before performing Random Forest analysis, the first step that must be done is to call the data to be analyzed into Python using the pd.read_csv command. Here is the data that was successfully called using Python.

	Merek	Series	Model No.	Material	Gender	Harga
0	Alexandre Christie	Ac-Asteria	ACF-8518	Stainless Steel	Pria	1105000
1	Alexandre Christie	Ac-Asteria	ACF-8521	Stainless Steel	Pria	910000
2	Alexandre Christie	Ac-Asteria	ACF-8602	Ion Plated	Wanita	1007000
3	Alexandre Christie	Ac-Asteria	ACF-8604	Stainless Steel With PVD Coating	Wanita	1007000
4	Alexandre Christie	Ac-Asteria	ACF-8612	Stainless Steel	Pria	1007000

Figure 4. Watch Data

After successfully recalling the data, the next step is to convert the categorical data into numeric so that analysis can be carried out at the next stage; the categorical data will be changed by converting it using the encoder label method. The variables to be converted to numeric are brand, series, model no., material, and gender variables. The following is the result of data conversion using encoder label.

	Merek	Series	Model No.	Material	Gender	Harga
0	2	0	252	10	0	1105000
1	2	0	254	10	0	910000
2	2	0	284	7	2	1007000
3	2	0	285	16	2	1007000
4	2	0	292	10	0	1007000
...	...	...	...	...	...	...
6123	9	139	1908	10	2	3816000
6124	9	139	1930	10	2	4240000
6125	9	139	1944	10	2	4928000
6126	6	147	558	0	0	1589000
6127	6	147	558	0	0	1413000

Figure 5. Watch Data Labeling

3.3.   Random Forest Analysis

Training data or training data is data used to test an algorithm while testing data or test data is data used to test and determine the performance of the model formed based on training data. The following is the proportion obtained from training data and testing data, with the proportion of training data being 0,8 and testing data 0,2.

Table 2. Training Data and Testing Data

| Description	Training Data	Testing Data	Total
Proportion	80%	20%	100%
Total	4902	1226	6128

Generally, training data has a larger share proportion than testing data. It can be seen in **Table 2** that the data is divided into two parts, namely training data and testing data, with a proportion of 80% for training and 20% for testing data taken randomly. So that the training data is 4902 data, while the testing data is 1226 data. After dividing the data, the next step is to determine the parameters by tuning to get the best parameters. The following is the result of parameter tuning.

Table 3. Parameter Tuning Result

Parameter	Randomized Search CV	Best Parameters
n_estimators	10; 100; 200; 400; 600; 800; 1000; 1200; 1400; 1600; 1800; 2000; 2200; 2400; 2600; 2800; 3000; 3200; 3400; 3600; 3800; 4000; 4200; 4400; 4600; 4800; 5000; 5200; 5400; 5600; 5800; 6000; 6200; 6400; 6600; 6800; 7000	3337
max_depth	10; 20; 30; 40; 50; 60; 70; 80; 90; 100; 110	100
max_features	auto; sqrt	Auto
min_samples_leaf	1; 2; 4	1
min_samples_split	2; 5; 10	10
bootstrap	True; False	False

Table 3 above is the result of tuning using a randomized search CV with random determination and cross-validation for three repetitions of each parameter. So that the best parameters are obtained, where the number of trees formed is 3337 trees, then for the maximum number of levels in each decision tree of 100 branches, the maximum number of features considered to separate a node is to use auto, then for the minimum number of data points in leaf nodes, which is allowed as much as 1, the minimum number of data points placed on a node before the node is split is 10, and the method of sampling data points with or without replacement is False.

After obtaining the best parameters, the next step is to make a comparison plot of the actual data and predictive data using the testing data that has been built previously. The following is a plot display of the actual data comparison with the predicted data from the watch data. It can be seen that the prediction graph has followed the graph of the actual data, which indicates that the predicted value obtained from the model has a value that is almost close to the actual value. So, from these results, it can be seen the influence between the independent variables on the dependent variable by using the right algorithm in order to get the best model with a high level of accuracy in order to produce a predictive value that is closer to the actual value.
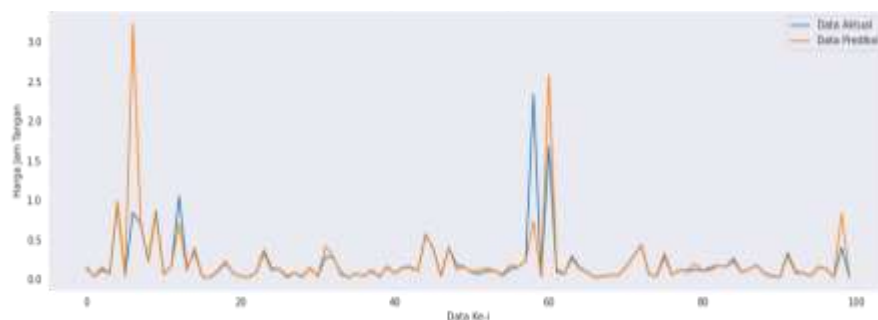
Figure 6. Plot Actual Data and Predicted Data

The following is a comparison of the actual data with the predicted data displayed in tabular form. It can be seen in **Table 4**, that the prediction data is almost close to the actual data or the predicted price obtained by the model is not much different from the original price of the watch data. For more details, see **Table 4** below.

Table 4. Actual Price and Predicted Price

The i-th Data	Actual Price (Rp)	Predicted Price (Rp)
1	1.105.000	1.083.111,111
2	910.000	1.083.111,111
3	1.007.000	1.083.111,111
4	1.007.000	1.083.111,111
…	…	…
6125	4.240.000	7.510.500
6126	4.928.000	7.510.500
6127	1.589.000	1.453.750
6128	1.413.000	1.453.750

The next step will be to evaluate the model by looking at the value of the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the level of accuracy using 3337 trees. It can be seen in the following table that the RMSE value is 2361962,2681 with a MAPE value of 15,02%. A data will get a high accuracy value if it gets a small RMSE value, and a model will produce the smallest possible error rate if a low MAPE value is obtained. So that the accuracy of the model formed is 84,98%. More clearly, see **Table 5**.

Table 5. Accuracy Results

RMSE	MAPE	Accuracy
2361962,2681	15,02%	84,98%

After evaluating the model by looking at the value of RMSE, MAPE, and the level of accuracy, then the factors that influence the price variable using features' importance will be seen. It can be seen in the following figure, that the most influential variable on the price variable is the material variable, then the brand variable, then the model no., series, and finally gender. More clearly, see the following figure.
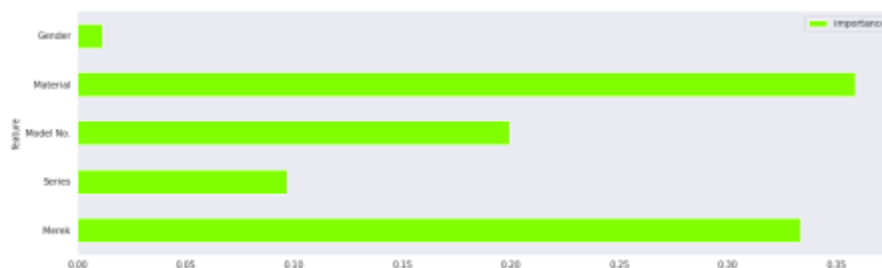
Figure 7. Features' Importance

To make it easier to see what factors affect the price variable, a table is formed, as shown in **Table 6** below. It can be seen that the variable that most influence the price of a watch is material with an importance level of 0,359, then the second is a brand with an importance level of 0,334, then model no with an importance level of 0,199, series with an importance level of 0,097, and gender with the level of importance is 0,012. More clearly, see **Table 6**.

Table 6. Features' Importance

Variable	Importance
Material	0,358872
Brand	0,333509
Model No.	0,199277
Series	0,096766
Gender	0,011575

After the model is formed and the factors that affect the price variable are known, the next model will be stored and will be used in the web design process using the pickle function in Python. Here's how to save the model and try to predict the price of one of the watches, where 1 = Alba, 49 = Fashion, 353 = AG200, 10 = Stainless Steel, 2 = Women. So that the results of the prediction of the price of watches for the Aigner brand are Rp. 603.666,67. For more details, see the following figure.

```
import pickle
pickle.dump(regressor, open('model.pkl','wb'))
model = pickle.load(open('model.pkl','rb'))
print(model.predict([[1,49,353,10,2]]))

[603666.6666667]
```

Figure 8. Prediction Model

3.4. Best Model

After conducting an analysis using the random forest method, the next step is to see whether the random forest model used in the analysis process is the best model when compared to other models, therefore four models will be compared to prove whether the random forest model is proven to be the best model. There are four models that will be tested to see the highest accuracy value, namely the Logistics Regression model, K-Nearest Neighbors Regression (KNN), Random Forest, and Support Vector Regression (SVR). The following are the accuracy results obtained from each model.

Table 7. Model Comparison

Model	Accuracy
Logistic Regression	0,224824
KNN	0,828687
Random Forest	0,858682
SVR	0,078414

It can be seen in **Table 7** that the Random Forest method is proven to be the best model in conducting the analysis, because it gets the highest accuracy value compared to other models, which is 0,858682 or 85%.

3.5. Creating a Prediction System Using Flask and Heroku

The following is a script used to create an HTML web display that is saved under the name "index.html", where later, the initial web page will contain the brand variable, series, model no., material, gender, and links from the watch website. Then it will be saved in the "templates" folder. The following script can be changed as desired, so that the display looks more attractive. Making scripts is done with the help of Visual Studio Code software. More clearly, see the following figure.

Figure 9. Script Index

After creating the initial web page, a script will then be created that connects Python with a web interface that was previously created using the flask function, so that the prediction results can be displayed online, with the following steps: filling out the watch price prediction form, then the form data will be

executed using the built-in model, then it will be returned in the form of prediction results which are displayed on the watch prediction web page. This script is saved under the name "app.py" and placed under the "templates" folder.

```python
import os
import numpy as np
import flask
import pickle
from flask import Flask, render_template, request

app = Flask(__name__, template_folder='templates')

@app.route('/')
def index():
    return render_template("index.html")
def ValuePredictor(to_predict_list):
    to_predict = np.array(to_predict_list).reshape(1,5)
    loaded_model = pickle.load(open("model.pkl","rb"))
    result = loaded_model.predict(to_predict)
    return round(result[0],2)

@app.route('/',methods = ['POST', 'GET'])
def result():
    if request.method == 'POST':
        to_predict_list = request.form.to_dict()
        to_predict_list=list(to_predict_list.values())
        to_predict_list = list(map(float, to_predict_list))
```

Figure 10. Script App

After finishing creating scripts for web pages and scripts that connect python to web displays, the next step is to run the script to see if the system that has been created is running successfully or not. The first step to run the system script is to open Visual Studio Code, then click the app.py file and click running, if successful, a link will appear to open the web as shown below.

```
* Restarting with watchdog (windowsapi)
* Debugger is active!
* Debugger PIN: 135-482-093
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Figure 11. Running App

After successfully running the script and no errors occur, the next step will be to deploy it to the internet so that it can be accessed by the wider community using the help of Heroku. The process of deploying to the internet using Heroku will be assisted by cmd. The first step is to open cmd, then type git init to create an empty repository and enter. The output will look like the following.

```
C:\Windows\System32\cmd.exe - git  push heroku master
Microsoft Windows [Version 10.0.19041.804]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\KIKI\Data TA>git init
Initialized empty Git repository in C:/Users/KIKI/Data TA/.git/
```

Figure 12. Add Empty Repository

Then type git add . to add all files and folders contained in the TA Data folder and enter, the output will look like the following.

```
C:\Users\KIKI\Data TA>git add .
warning: LF will be replaced by CRLF in .ipynb_checkpoints/Perbandingan Model Terbaik-checkpoint.ipynb.
The file will have its original line endings in your working directory
warning: LF will be replaced by CRLF in .ipynb_checkpoints/Random Forest TA-checkpoint.ipynb.
The file will have its original line endings in your working directory
warning: LF will be replaced by CRLF in Perbandingan Model Terbaik.ipynb.
The file will have its original line endings in your working directory
warning: LF will be replaced by CRLF in Random Forest TA.ipynb.
The file will have its original line endings in your working directory
warning: LF will be replaced by CRLF in env/Lib/site-packages/easy_install.py.
The file will have its original line endings in your working directory
warning: LF will be replaced by CRLF in env/Lib/site-packages/pip-20.2.3.dist-info/INSTALLER.
```

Figure 13. Add Files and Folders

Then type git status to see the location of the files and folders that will be included in the master as follows.

Figure 14. View File and Folder Location

After the process of entering files and folders is complete, a new folder will automatically appear, namely the "git" folder in the TA Data folder as follows.

Figure 15. Git Folder

Then login to Heroku via cmd by typing Heroku Login as follows.

Figure 16. Login Heroku

Then Heroku will automatically open in the browser as shown in the following image, then click login.

Figure 17. Heroku Login Page

The next step is to create a web name by typing heroku create prediksijamtangan. After successfully creating a web name, next type git remote –v to see if git with heroku has been connected or not. If it is connected, it will produce output like the following.

Figure 18. Git Remote

Then it will be deployed to heroku by typing git push heroku master in cmd as follows.

Figure 19. Deploy Heroku

After all the processes are complete, the watch prediction web is ready to be used online by the whole community by opening the watch prediction web using the link that has been given previously, namely https://predikjamtangan.herokuapp.com/. The following is a watch price prediction web display, where there are already 5 form fields that must be filled in to find out the price of the watch we want to predict, namely the brand, series, model no., material, and gender columns, and at the end, there is a link to access watch product website, if you want to buy a watch.

Figure 20. Watch Prediction Website

4. CONCLUSIONS

A conclusion should give a summary of the following:

1. Based on the percentage of watch product availability by brand on the machtwatch website, it was found that the most sold products were products with the Casio brand with a percentage of product availability of 36% or as many as 2207 products, while the least sold products were products with the Citizen brand with a percentage of 1% or as many as 74 products. Then for the availability of watch types based on movement, style, and needs on the machtwatch website, it was found that the most widely sold types of watches on the website are watches with Quartz movements with a total availability of 1357 products, while the Pilot watches are types of watches, which is sold at least with a total of 3 products. Then when viewed from the average selling price of watches by brand, it was found that watches with the Tissot brand had the highest average watch price of Rp.9.698.712,- while the lowest average watch price was found in products with the Alba brand Rp.724.847,-

2. After forming the best model using the Random Forest analysis method with a total of 3337 trees, an accuracy rate of 84,98% is obtained.

3. The variable that most influence the price of watches is the material variable with an importance level of 0,359.

4. The prediction web system that has been formed consists of 5 form fields that must be filled in to find out the price of the watch you want to predict, namely the column brand, series, model no., material, and gender, and at the end, there is a link to access the watch product website. Wrist, if we want to buy a watch and can be accessed via the link that has been included. To access the watch price prediction web, we can click the link https://prediksijamtangan.herokuapp.com/.

REFERENCES

[1] A. W. Pradipta and B. T. Indrojarwo, "Desain Jam Tangan Kayu dengan Konsep Jujur Material dan Inklusif," *Jurnal Sains dan Seni ITS,* vol. 5, no. 2, pp. 243-248, 2016.

[2] K. Saputra, "Pengaruh Gaya Hidup Citra Merek dan Negara Asal terhadap Sikap Mahasiswa Universitas Andalas Dalam Keputusan Pembelian Produk Jam Tangan Casio," Padang, 2019.

[3] L. Hakim, "Pandemi Covid-19, Penjualan Jam Tangan Garmin Naik Hingga 50 Persen," SindoNews, 19 December 2020. [Online]. Available: https://daerah.sindonews.com/read/273210/704/pandemi-covid-19-penjualan-jam-tangan-garmin-naik-hingga-50-persen-1608293531. [Accessed 05 January 2022].

[4] S. Arimasen, "Meski Harga Murah, Ini 7 Alasan Hindari Jam Tangan KW," Jamtangan.com, 04 August 2020. [Online]. Available: https://blog.jamtangan.com/meski-harga-murah-ini-7-alasan-hindari-jam-tangan-kw/. [Accessed 26 December 2021].

[5] Majalah Radatime, "Kerugian Membeli Jam Tangan KW/Palsu," 26 Juni 2021. [Online]. Available: https://majalah.radatime.co.id/kerugian-membeli-jam-tangan-kw-palsu/. [Accessed 26 December 2021].

[6] T. Purwa, "Perbandingan Metode Regresi Logistik dan Random Forest Untuk Klasifikasi Data Imbalanced," *Jurnal Matematika, Statistika dan Komputasi,* vol. 16, no. 1, pp. 58-73, July 2019.

[7] Y. A. Tampil and d. , "Analisis Regresi Logistik untuk Menenukan Faktor-Faktor yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *JdC,* vol. 6, no. 2, pp. 56-62, 2017.

[8] M. Nanja and P. , "Metode K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Harga Komoditi Lada," *Jurnal Pseudocode,* vol. 2, no. 1, pp. 53-64, February 2015.

[9] R. Amanda, H. Yasin and A. Prahutama, "Analisis Support Vector Regression (SVR) dalam Memprediksi Kurs Rupiah Terhadap Dollar Amerika Serikat," *Jurnal Gaussian,* vol. 3, no. 4, pp. 849-857, 2014.

[10] B. Scholkopf and A. J. Smola, Learning With Kernels, London, England: The MIT Press, 2002.

[11] S. A. Zega, "Penggunaan Pohon Keputusan Untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Yogyakarta, 2014.

[12] R. Hartanto, "Penerapan Algoritma C 4.5 dengan Menggunakan Metode Decision Tree untuk Memprediksi Target Produksi Casting Part PT. Eagle Industry Indonesia," Bekasi, 2018.

[13] M. Dhawangkhara, "Prediksi Intensitas Hujan Kota Surabaya dengan Matlab Menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya)," Surabaya, 2016.

[14] N. Sintyaningrum, "Pemilihan Input dengan Random Forest Pada Model Time Series Regression dan Double Seasonal ARIMA Untuk Peramalan Data Pemakaian Beban Listrik Jangka Pendek," Surabaya, 2017.

[15] B. Prasojo and E. Haryatmi, "Analisis Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi,* vol. 7, no. 2, pp. 79-89, 2 September 2021.

[16] R. A. Haristu, "Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground," Yogyakarta, 2019.

[17] D. D. Ayani, H. S. Pratiwi and H. Muhardi, "Implementasi Web Scraping untuk Pengambilan Data Pada Situs Marketplace," *Jurnal Sistem dan Teknologi Informasi,* vol. 7, no. 4, pp. 257-262, October 2019.