Classification of Historical Anatolian Coins with Machine Learning Algorithms

Ramazan Ünlü, Ph.D. *   iD

Assist. Prof., Department of Management Information Systems, Faculty of Economic and Administrative Sciences, Gumushane University,
Gumushane, Turkey, ramazanunlu@gmail.com

* Gümüşhane Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Gümüşhanevi Kampüsü, Bağlarbaşı Mahallesi, 29100, Merkez Gümüşhane Türkiye

ABSTRACT	To find out which period the historical coins belong to requires a number of scientific procedures that archaeologists or experts can do. These operations can often be time-consuming and demanding operations. From this point on, in this study, the automatically classification of historical coins by using machine learning methods is discussed. Being able to use machine learning methods to classify historical coins can help experts and can become an analysis tool without the need for scientific tests for non-experts. For this purpose, some physical properties of different coins used in Anatolian geography were collected and classified by various machine learning methods named SVM, Random Forest, Bagging, and Decision Trees. Also, two different missing values strategies are deployed in conjunction with each chosen method. Based on our findings, random forest method together with imputing missing values with mean gives an acceptable results with the accuracy rate of %71, although there are some limitations such as high rate of missing values and working with a small dataset.
Keywords:	Historical Coins, Machine Learning, SVM, Decision Trees, Bagging, Random Forest, Classification

1. Introduction

Machine learning, dating back to the 1950s, from the discovery of perceptron to the present day, has been deeply studied by many researchers and developed much more complex methods and many of them used in the different fields. In recent years, many different applications of machine learning methods such as image processing (Long, Shelhamer, & Darrell, 2015), natural language processing (Collobert et al., 2011), text mining (Feldman & Sanger, 2007), autonomous vehicles (Sallab, Abdou, Perot, & Yogamani, 2017) have been effectively used in various fields such as medicine (Deo, 2015; Obermeyer & Emanuel, 2016), manufacturing (Priore, De La Fuente, Gomez, & Puente, 2001; Wuest, Weimer, Irgens, & Thoben, 2016), finance (Bose & Mahapatra, 2001; Trippi, By-Lee, & Jae, 1995), and education (Baker, 2010; Zhu, 2015). Especially in the last decade, the development of technology, the increase in computer power and the ability to store data has made machine learning quite popular. When we look at the working mechanisms of machine learning methods, what we will see is that a machine can solve very complex problems in a much shorter time when the human mind will be inadequate. As this fact may have some drawbacks (e.g. reduction of mankind's labor need in the future), there are also some advantages (e.g. the ability of machines to diagnose before and more accurately than doctors). From a more technical point of view, we can aggregate machine learning methods in roughly three topics: Supervised Learning, Unsupervised learning, and Semi-Supervised learning. To understand these concepts more deeply, we need to look at what data means. If we know the correct output of data samples in a dataset, this dataset is called labelled data. If the correct output is unknown, this dataset is called unlabeled data. Generally, supervised learning methods are used for labelled data, while unsupervised methods are used for unlabeled data. However, if the correct output of some data samples is known, some are unknown, then semi-supervised methods can be used.

Generally, machine learning methods are used to classify quickly and effectively in supervised learning problems. From this point on, we examined whether machine learning based methods can be utilized to classify historical coins found in Turkey. The age of a historical thing is usually determined by archaeologists by applying various methods. Some of those techniques are listed below (Michels, 1973).

- Radiocarbon Calibration

- Geomorphological Dating

- Relative Dating

- Nautical Applications of Dendrochronology

- Art-historical Dating

- Dendroclimatology

Many of these techniques require various materials, chemical materials and, most importantly, a specialist. In addition, a certain extent of time is spent in order to efficiently implement each technique. Thus, the calculation of the age of a historical object can be both laborious, time-consuming and costly. And these handicaps are the starting point of the origin of main question of this study which is that "Can be the period of a historical coin predicted by machine learning methods?"

Unfortunately, the classification of historical coins using machine learning methods is very little in the literature. Especially when we look at the studies that focus on the classification of historical coins found in Turkey, we see that no work has been done. Therefore, this study is the first in terms of the data used and the concept created.

However, there are some exceptional studies that use different methods to classify coins. These studies solely focus on the classification of modern coins. Maaten and Boon developed a system called COIN-O-MATIC using images of various coins. This system consists of five phases: segmentation, feature extraction, preselection, classification, and verification (L. J. Van Der Maaten & Poon, 2006). They have used nearest-neighbor approach for the classification purpose and have tested the model accuracy in verification stage. They have achieved the 67.31% accuracy on a dataset which consists of coins at MUSCLE CIS Coin Competition (Nölle, Rubik, & Hanbury, 2006). COIN-O-MATIC system differs in the characteristics of data and coins used from our work. While Maaten and Boon used pictures of coins, we used some physical properties of coins, such as diameter, weight, etc. Also instead of historical coins, they trained their system on a set of modern European coins that were gathered after the euro released. (Fukumi, Omatu, Takeda, & Kosaka, 1992) have used a rotation-invariant neural pattern recognition system to classify the pictures of 500 yen and won coins. The Eigen space method used by (Huber, Ramoser, Mayer, Penz, & Rubik, 2005) is an another study in this field. They have used 11949 different modern coin to test their method and achieved %93.23 accuracy rate. (Davidsson, 1997) have used decision tress to classify Canadian and Hong Kong modern coins. Decision tree model is fed by five different attributes diameter, thickness, conductivity1, conductivity2, and permeability and achieved more than %90 accuracy with several transformation of decision trees. Edge detection and gradient directions (Reisert, Ronneberger, & Burkhardt, 2006; L. Van Der Maaten & Postma, 2006) and contour and text features (Nölle et al., 2003) are also the used methodologies can be found in the literature.

Most of these studies are often studies on the classification of modern coins. In addition, almost all of these works have used images of coins as raw data. On the other hand, historical coins are used, but their number is very limited. In the study of (Zambanini & Kampel, 2011), ancient coins are classified based on coins' image. They have used SIFT flow method in order to find the similarity of coin images. The dataset contains 24 classes of early Roman Republican coinage and they have achieved a rate of classification accuracy 74%.

None of these studies focuses on prediction of the era of the historical coins, but they have aimed the find what kind of coin is that. Also, considering the coins used in the Anatolian geography in the past, none of those studies in the literature has not rerouted to a forecasting model by using those anatolian coins. In this regard, our work is the first study to offer a forecasting model for used coins in Anatolia and for determining era of historical coins. Also as different from previous studies, some physical properties of coins were used as the feature of the training dataset, rather than pictures of coins.

The main objective of the proposed study is testing various machine learning methods and strategies to classify coins used in the aforementioned geography. As will be explained in detail in the future sections, even if we have some limitations

because of the data, various algorithms and strategies have been tried to compare each strategies from different aspects.

The study is constructed as follows; In section 2, we have described the methodology of used methods, in section 3 we give the detailed results of algorithms in conjunction with some strategies deployed to deal with the missing values. Finally, we have made the conclusion and discussed the performance of algorithms and possible future studies.

2. Methodology

The successful results of machine learning methods for an estimation problem is fundamentally dependent on three things; 1) Method Used 2) Collected data and 3) Data preprocessing. For this reason, different methods were compared with different preprocessing strategies. Also some of the disadvantages of the dataset are described.

2.1. Dataset

The dataset used in this study contains some physical properties of various historical coins. A total of 100 different coin information is available, and these coins are known to be from 1200 to 1900. A few examples of these coins are shown in the *Figure 1* below.

Figure 1. Real pictures of some historical coins.

The dataset has four main attributes which is used in prediction model. These are alloy, shape, weight and diameter. Our first constraint of the dataset arises at this point. Unfortunately, there are only four features mentioned above for the dataset. There are also a large number of unknown values in some properties. And that forced us to implement different strategies before we set up an estimation model.

Therefore, each chosen method has been used and compared with different strategies. The Table 1 below shows the top ten examples of the raw dataset

Alloy	Shape	Weight (gr)	Diameter (mm)	Release date
Bronze	Formless	10.9	32	1199
Silver	Formless	2.6	20	1237
Silver	Formless	1.12	19	1326
Silver	Round	1.3	13	1362
Silver	Formless	1.15	18	1362
Silver	Formless	0.9	10	1389
Silver	Formless	1.16	15	1389
Copper	Formless	1.33	18	1389
Silver	Formless	1.2	11	1411
Copper	Round	0.6	9	1413
Silver	Formless	1.15	11	1421
Silver	Round	1.3	12	1421
Silver	Formless	0.93	12	1451
Silver	Formless	0.98	11	1451
Bronze	Formless	1.05	12	1451

Table 1. Samples from raw data.

As shown in Table 1, alloy and shape properties are categorical variables and there is no superiority among themselves. For example, the fact that a coin has a round shape does not mean that it will be mathematically more weighted in the prediction model than a coin with an uncertain shape. Therefore, dummy variables for these properties were created and the dataset was converted to the same as shown in the Table 2 below.

Alloy1	Alloy2	Alloy3	Alloy4	Alloy5	Alloy6	Shape1	Shape2	Shape3	WEIGHT(gr)	DIAMETER (mm)
1	0	0	0	0	0	1	0	0	10.9	32
0	1	0	0	0	0	1	0	0	2.6	20
0	1	0	0	0	0	1	0	0	1.12	19
0	1	0	0	0	0	0	1	0	1.3	13
0	1	0	0	0	0	1	0	0	1.15	18
0	1	0	0	0	0	1	0	0	0.9	10
0	1	0	0	0	0	1	0	0	1.16	15
0	0	1	0	0	0	1	0	0	1.33	18
0	1	0	0	0	0	1	0	0	1.2	11

Table 2. The data after creating dummy variables.

Finally, the appropriate labels for each data sample are determined based on the release date of the coin. At this stage, two different strategies are implemented. In the first strategy, each century is considered as a class, in the second strategy each two centuries is considered as a class. As a result of the first and second strategy, four and seven different classes were obtained respectively.

2.2.   Algorithms

2.2.1.   Support Vector Machines

Support vector machines (SVM) were first proposed by Vapnik and Cortes for binary classification problems in 1995 (Cortes & Vapnik, 1995). The SVM binary classification

method was then extended for both regression and multiple classification problems (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Quinlan, 1986). To mathematically make it more concrete let say that we have a training dataset $\{(x_1, y_2), \ldots, (x_l, y_l)\}$, where each $x_i \in R$ the decision function is given by Equation 1.

$$f(x) = w\phi(x) + b \tag{1}$$

With respect to $w_i \in R$ and $b \in R$, where ϕ denotes a non-linear mapping transformation from R^n to higher dimensional space. To assure $f(x)$ is as flat as possible, we require to solve it with the minimal norm value as shown in Equation 2.

$$J(w) = \frac{1}{2}\|w\| \tag{2}$$

Subject to all residuals which have a value less than ε; or, in equation form (see Equation 3):

$$w\phi(x_i) + b - y_i \leq \varepsilon \tag{3}$$

We can infer that it is not possible to meet this condition for the all points. So, we can add two different slack variables ξ^+ and ξ^- to make some flexibility and edit the formula as shown in Equation 4:

$$J(w) = \frac{1}{2}\|w\| + C\sum_i^n \xi^+ + \xi^- \tag{4}$$

Subject to:

$$y_i - (w\phi(x_i) + b) \leq \varepsilon + \xi^+$$

$$(w\phi(x_i) + b) - y_i \leq \varepsilon + \xi^-$$

$$\xi^+ \geq 0$$

$$\xi^- \geq 0$$

where C is a constant value that deal with the penalty values which enforce the variable which lies outside the ε margin and help to avoid being overfitting.

Finally, we can calculate the loss function that ignores the error if the predicted value is less than or equal to margin of ε. Therefore, it can be calculated based on Equation 5.

$$f(x) = \begin{cases} 0, & if\ w\phi(x_i) + b - y_i \leq \varepsilon \\ |w\phi(x_i) + b - y_i| - \varepsilon, & otherwise \end{cases} \tag{5}$$

The optimization problem given above can be solved in dual form for the mathematical convenience.

2.2.2. Decision Trees

Decision trees are one of the machine learning methods commonly used in classification and regression problems (Quinlan, 1986). The basic principle of decision trees is to divide data under various conditions and produce an estimate at the final stage. The following Figure 2 shows an example of a decision tree. The node at the top of a decision tree is called root node and can be considered the first condition that divides data. Then various conditions are added and this process is continued until a final result is obtained. Here are some terms for decision trees.

- **Root Node:** Root node is the node located at the top of the decision tree and represents the entire population. So the whole dataset starts to split from here for the first time

- **Splitting:** Splitting is the term used for dividing a node into two or more sub nodes.

- **Decision Node:** It is the name of divided node.

- **Leaf/ Terminal Node:** If a node is no longer divisible, it is named as leaf or terminal node.

- **Pruning:** Pruning is used to reduce the size of tree by removing some nodes.

- **Branch / Sub-Tree:** A sub section of entire tree is named branch or sub-tree.

- **Parent and Child Node:** When a node is divided, it is called the parent of the new created sub-nodes. The newly formed sub-nodes are called children of parent node.
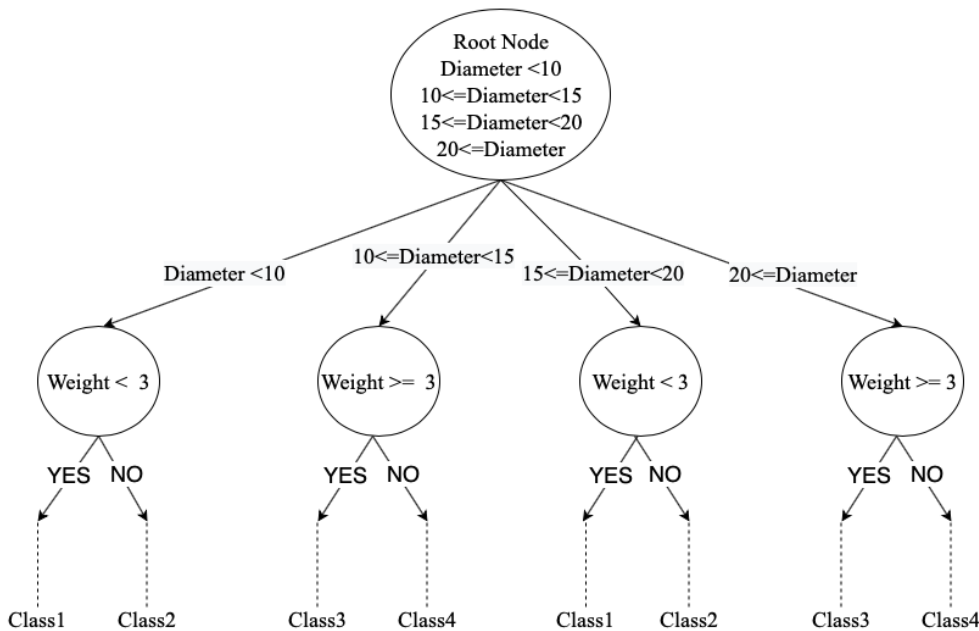
Figure 2. A Toy Example of Decision Trees Structure.

2.2.3. Random Forest

Although decision trees produce good results in many prediction problems, using a shallow decision tree can poorly perform, while a very deep decision tree causes overfitting problem. To eliminate these problems, researchers have developed the random forest method (Breiman, 2001). While Random Forest's basic logic is the same as decision trees, the random forest method uses many decision trees to find the best prediction by deploying majority voting method. The principal of a random forest model is given in the following Figure 3.
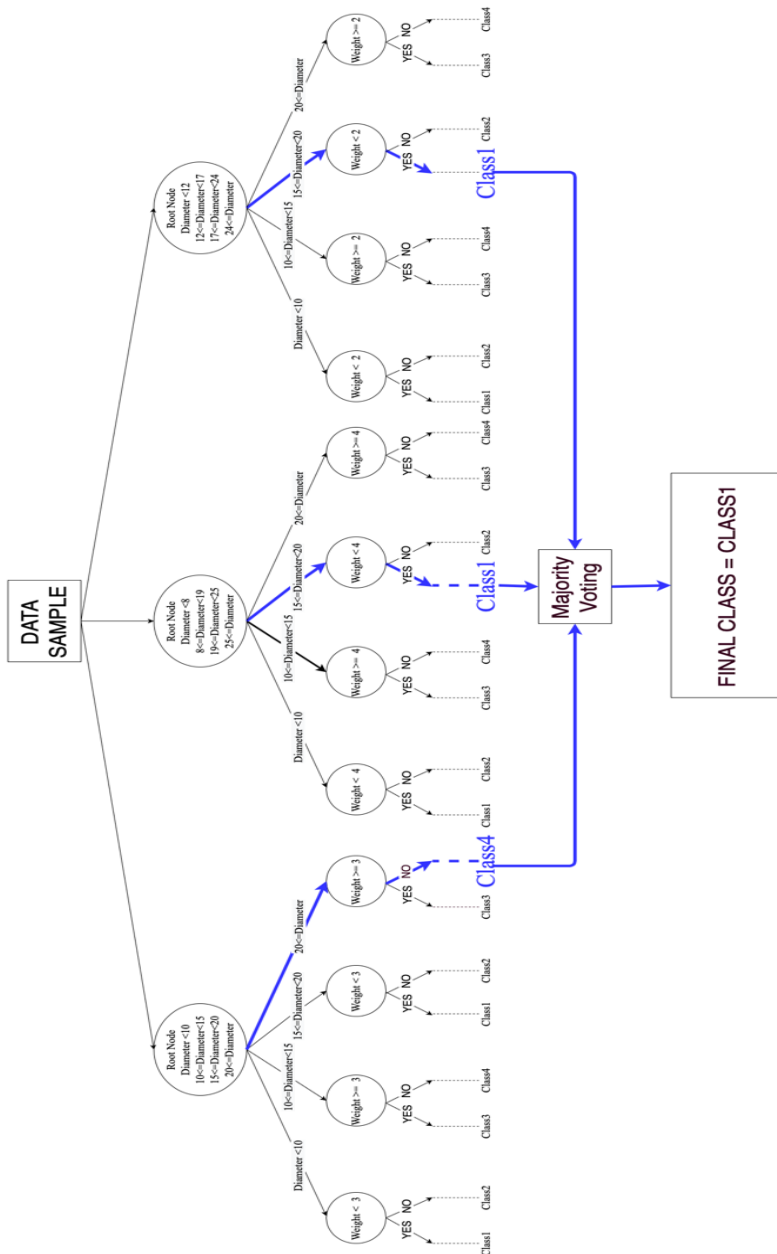
Figure 3. A Toy Example of Random Forest Structure.

2.2.4. Bagging

The Bagging method has been developed to achieve more successful results than a single machine learning method by combining several individual machine learning methods (Breiman, 1996). In this method, each chosen classifier is applied to a group of different samples randomly selected from the raw dataset Initially, different samples are created from a dataset, i.e. a different new dataset. Then any machine learning method (e.g. SVM) is applied to each dataset created, resulting in different prediction models. At the last stage, these models are tested with previously unused data, and the final estimate is based on the average or majority of the results obtained from these models. The following Figure 4 shows the working principle of the bagging method.
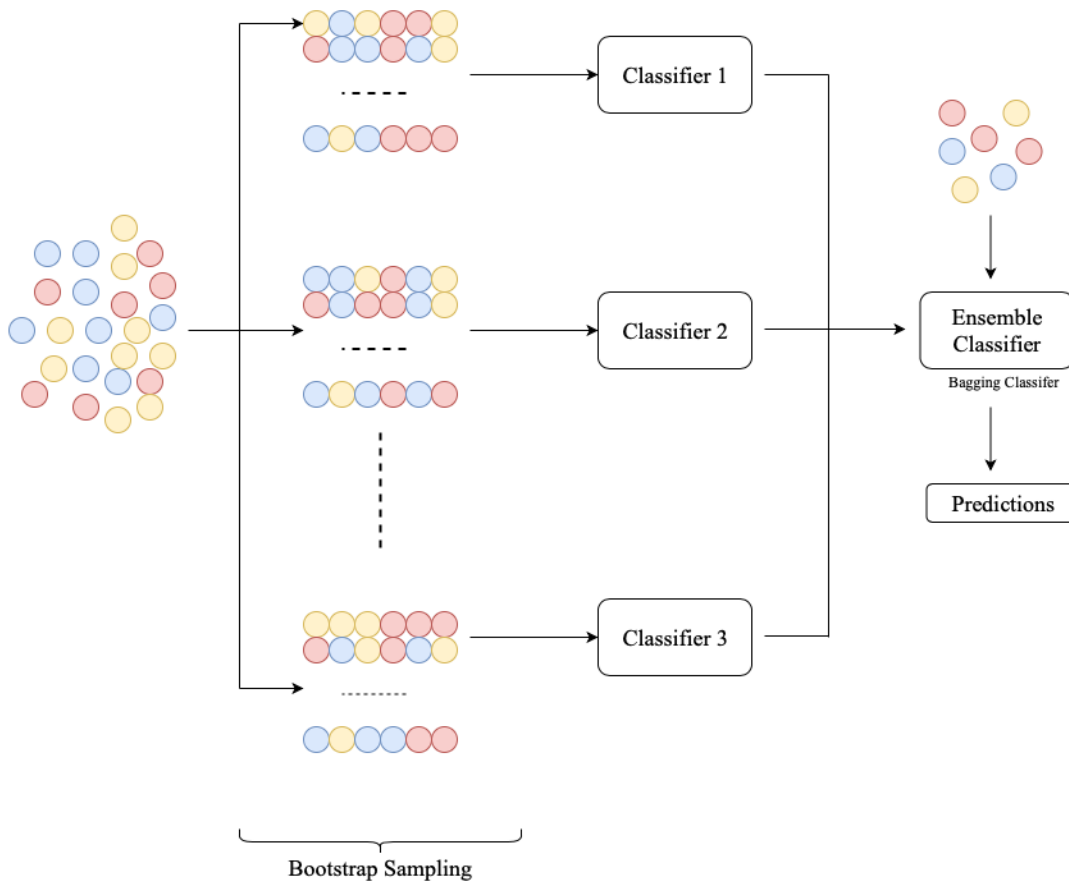
Figure 4. The Structure of Bagging Methodology

2.3.   Missing Values Strategies

As already mentioned, some properties in the dataset have unknown values. For example, the weights of some coins are unknown. Two different methods of imputation were used to eliminate these problems. The first method replaces the unknown values with the mean of known values and the second method replaces the median of known values with unknown values. Those strategies are simply described below.

2.3.1.   Imputing missing values with mean

Let's say 90 out of 100 samples are known, and 10 of them are not known. In this case, the known 90 values are averaged and used as the actual value of 10 unknown values when creating the estimation model. This method has its advantages, as well as its drawbacks. For example, when the sample with the unknown value is not excluded from the dataset, important information can be saved, but the estimation model can be misdirected if extreme values are present in known values.

2.3.2.   Imputing missing values with median

Impute unknown values with median works on the same principle as impute unknown values with mean. The only difference is to use median instead of average. If the

known values do not have extreme values, this method can produce more successful results than the other method.

3. Results

In this section, we summarized the results of four different machine learning algorithms used with two different missing values strategies and two different labeling process. At the end, various performance criteria of 16 different strategies were looked at. Each method has been tried to achieve the most reliable results by running the 10-fold cross-validation method.

The first algorithm we have investigated is the SVM. Table 3 shows the results of SVM model in terms of four different evaluation metrics.

Strategies	Accuracy	Precision	Recall	F-Measure
Class number=4 Imputing missing values with mean	0.65	0.715	0.65	0.61
Class number=4 Imputing missing values with median	0.64	0.722	0.66	0.627
Class number=7 Imputing missing values with mean	0.4	0.304	0.4	0.315
Class number=7 Imputing missing values with median	0.39	0.293	0.39	0.309

Table 3. The results of SVM model

According to SVM results, the most remarkable point is that the strategy applied when creating labels seriously affects algorithm performance. However, SVM cannot be said to be very successful in both strategies. SVM caught the highest accuracy in a 4-class strategy where unknown values were imputed with mean.

Strategies	Accuracy	Precision	Recall	F-Measure
Class number=4 Imputing missing values with mean	0.69	0.704	0.69	0.686
Class number=4 Imputing missing values with median	0.69	0.704	0.69	0.686
Class number=7 Imputing missing values with mean	0.54	0.535	0.54	0.527
Class number=7 Imputing missing values with median	0.51	0.501	0.51	0.493

Table 4. The results of Bagging model.

The Bagging method yielded more successful results in all strategies compared to the SVM method with respect to the accuracy ratio assessment criteria. The most successful performance was when the number of classes was four, with 69% accuracy. On the other hand, when the number of classes was four, two different strategies applied for missing values had no effect on the performance of the model. The bagging method also demonstrated the same successful performance for three other evaluation criteria**.**

Strategies	Accuracy	Precision	Recall	F-Measure
Class number=4 Imputing missing values with mean	0.68	0.675	0.68	0.676
Class number=4 Imputing missing values with median	0.62	0.62	0.62	0.617
Class number=7 Imputing missing values with mean	0.45	0.442	0.45	0.444
Class number=7 Imputing missing values with median	0.5	0.488	0.5	0.489

Table 5. The results of Decision Tree Model

The most successful performance of the standard decision trees method was achieved with 68% accuracy in the strategy where the number of classes was four and missing values imputed with mean. Unlike the Bagging model, the missing values strategy has a relatively impact on the performance of the decision tree method.

Strategies	Accuracy	Precision	Recall	F-Measure
Class number=4 Imputing missing values with mean	0.71	0.723	0.71	0.706
Class number=4 Imputing missing values with median	0.69	0.704	0.69	0.686
Class number=7 Imputing missing values with mean	0.52	0.517	0.52	0.514
Class number=7 Imputing missing values with median	0.52	0.519	0.52	0.509

Table 6. The results of Random Forest Model.

The last method, Random forest, has given the highest accuracy compared to other methods. Although it was not able to produce the desired result when the number of classes was seven, it gives relatively acceptable results when the number of classes was four. The Random forest method reached 71% accuracy in the strategy where the number of classes was four and missing values were imputed with mean.

Possible causes and results of the findings and conditions that may be necessary for machine learning methods to give better results in the problem of classification of historical coins discussed in more detail in the Section 4.

4. Conclusion and Discussion

This study examined whether machine learning methods can determine the periods of historical coins. For this purpose, various information from 100 different historical coins have been collected. The collected data was analyzed using four different machine learning methods and the results were compared based on four different evaluation criteria as shown in Table 3, Table 4, Table 5 and Table 6. Random forest method with 71% accuracy showed the most successful performance in general.

However, when we look at the overall performance of the methods used, there were no satisfactory results according to the what generally excepted from a machine learning method. By investigating the collected dataset, we can clearly see the reasons for this. Unfortunately, the number of data we can collect for use in the study is very limited. Because only one hundred data can be collected from an open source, there may not be enough information to train algorithms. This may be one of the most likely reasons of algorithms did not perform much better.

Another limitation is that the dataset has more missing values than normal. 27 out of 100 coins do not have the information of diameter of the coins. A 27% loss of information could be another factor that could affect the performance of algorithms. In addition, the weight of 2 coins is not available. However, a loss of 2% information can be considered within an acceptable limit.

The Figure 5 below shows how lost values can affect data integrity. Considering that the data is sorted by release dates in this chart, we can see that many lost values are sequential or close to each other. This indicates that the majority of missing data belongs to coins in the same class. In this case, any machine learning algorithm can assign coins for that class to another class whose diameter average is close to the overall diameter average.
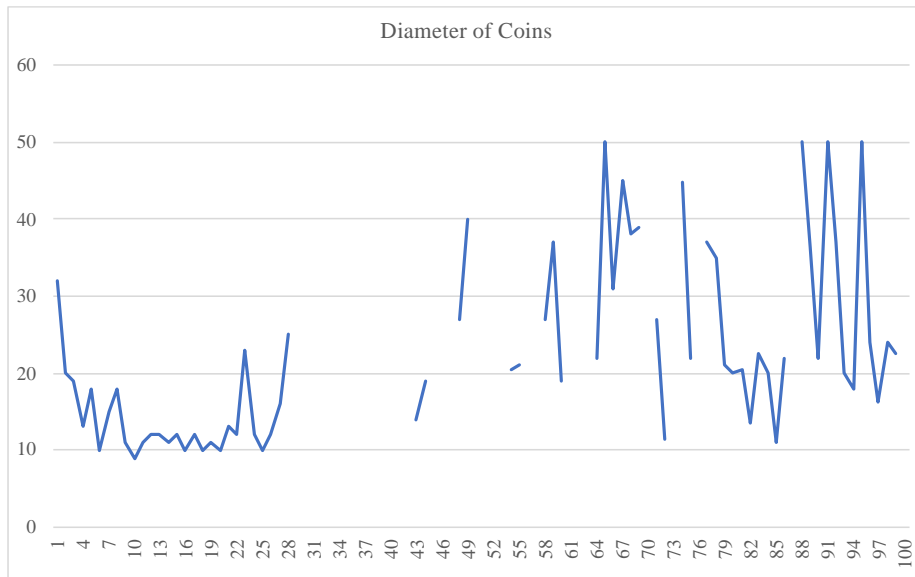
Figure 5. Diameter of coins. Coins are sorted by release dates

This can be made more concrete in the following way. The following Table 7 shows the confusion matrix of the random forest method. According to this matrix, the Random Forest method most misclassified the data samples belonging to class two and class three. 11 data samples belonging to class two and nine data samples belonging to class three are assigned to the wrong classes. The total number of data for these classes is 39 for class two and 38 for class three, respectively. Finally, when we look at missing data, all samples containing lost data belong to classes two and three. So we can say that being the majority of the lost data in a single class greatly affects the performance of the algorithm.

A	B	C	D	← Classified as
12	4	1	0	A=1
3	28	8	0	B=2
1	8	29	0	C=3
0	0	4	2	D=4

Table 7. Confusion Matrix of Random Forest Method

As mentioned above, another reason why the performance of algorithms is not at the desired level may be a small number of data. Especially, if there are very small number of data in one of the classes, in other words if the data is imbalanced. Assigning minority data instances to the right class can be even more difficult. We can embody this example by looking at the samples belonging to class four in the Table 7. Of 100 data samples, only six of them are in class four. Therefore, it is more difficult for the algorithm to learn the characteristic information of this class of data. In this regard, the imbalance data can be considered another reason why algorithm performance is not what should be expected.

Despite all the above-mentioned limitations, this study shows that machine learning methods can be used to predict periods of historical coins. If we take the Random Forest method that gives the best results, we see that the algorithm predicts correctly in 71 out of 100 coins, despite the serious limitations. This ratio is a sufficiently satisfying result as an initial work. In future studies, we believe that even more successful results can be achieved by increasing the number of samples, reducing the missing value ratio, and adding more features (e.g. thickness of coins).

In this way, a prediction can be made using artificial intelligence algorithms without the need for a specialist at least in the initial stage of tests for which used finding out which period the historical coins belong to.

References

Baker, R. (2010). Data mining for education. International Encyclopedia of Education, 7(3), 112–118.

Bose, I., & Mahapatra, R. K. (2001). Business data mining—a machine learning perspective. Information & Management, 39(3), 211–225.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12(Aug), 2493–2537.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Davidsson, P. (1997). Coin classification using a novel technique for learning characteristic decision trees by controlling the degree of generalization. In Ninth International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (pp. 403–412).

Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In Advances in neural information processing systems (pp. 155–161).

Feldman, R., & Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press.

Fukumi, M., Omatu, S., Takeda, F., & Kosaka, T. (1992). Rotation-invariant neural pattern recognition system with application to coin recognition. IEEE Transactions on Neural Networks, 3(2), 272–279.

Huber, R., Ramoser, H., Mayer, K., Penz, H., & Rubik, M. (2005). Classification of coins using an eigenspace approach. Pattern Recognition Letters, 26(1), 61–75.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431–3440).

Michels, J. W. (1973). Dating methods in archaeology.

Nölle, M., Penz, H., Rubik, M., Mayer, K., Holländer, I., & Granec, R. (2003). Dagobert-a new coin recognition and sorting system. In Proceedings of the 7th Internation Conference on Digital Image Computing-Techniques and Applications (DICTA'03), Syndney, Australia.

Nölle, M., Rubik, M., & Hanbury, A. (2006). Results of the muscle cis coin competition 2006. In Proceedings of the Muscle CIS Coin Competition Workshop, Berlin, Germany (pp. 1–5). Citeseer.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216.

Priore, P., De La Fuente, D., Gomez, A., & Puente, J. (2001). A review of machine learning in dynamic scheduling of flexible manufacturing systems. Ai Edam, 15(3), 251–263.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.

Reisert, M., Ronneberger, O., & Burkhardt, H. (2006). An efficient gradient based registration technique for coin recognition. In Proc. of the Muscle CIS Coin Competition Workshop, Berlin, Germany (pp. 19–31).

Sallab, A. E. L., Abdou, M., Perot, E., & Yogamani, S. (2017). Deep reinforcement learning framework for autonomous driving. Electronic Imaging, 2017(19), 70–76.

Trippi, R. R., By-Lee, P., & Jae, K. (1995). Artificial intelligence in finance and investing: state-of-the-art technologies for securities selection and portfolio management. McGraw-Hill, Inc.

Van Der Maaten, L. J., & Poon, P. (2006). Coin-o-matic: A fast system for reliable coin classification. In Proc. of the Muscle CIS Coin Competition Workshop, Berlin, Germany (pp. 7–18).

Van Der Maaten, L., & Postma, E. (2006). Towards automatic coin classification. na.

Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research, 4(1), 23–45.

Zambanini, S., & Kampel, M. (2011). Automatic coin classification by image matching. In Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage (pp. 65–72). Eurographics Association.

Zhu, X. (2015). Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In Twenty-Ninth AAAI Conference on Artificial Intelligence.