


Machine Learning Techniques For Analyzing Inscriptions From Israel

Daiki Tagami <dt2631_at_columbia_dot_edu>, Columbia University  <https://orcid.org/0000-0002-0923-1070>
Michael Satlow <michael_satlow_at_brown_dot_edu>, Brown University

Abstract

The date of artifacts is an important factor for scholars to get a further understanding of culture and society of the past. However, many artifacts are damaged over time, and we can often only get fragments of information regarding the original artifact. Here, we use the inscription data from Israel as a model dataset and compare the performances of eleven commonly used regression models. We find that the random forest model would be the optimal machine learning model to predict the year of inscriptions from tabular data. We further show how we can make interpretations from the machine learning prediction model through a variance important plot. This research shows an overview of how machine learning techniques could be used to resolve digital humanities problems by using the Inscription of Israel/Palestine dataset as a model dataset

1. Introduction

The study of antiquity is full of missing data. The evidence that does survive – whether texts on papyrus or parchment; inscriptions; coins; or archaeological – frequently survives only in damaged form. That problem, however, is compounded by two additional complications. First, many of these data have been unearthed in non-controlled excavations and have taken winding paths to libraries, museums, and the hands of private collectors, along the way losing valuable contextual information. Second, scholars have used a bewildering array of conflicting and often inherently vague reporting methods. My “Roman period,” for example, might be your “Byzantine period.” As a result of this situation, scholars in ancient studies frequently find themselves unable to place or date evidence that could be critical to our deeper understanding. Given the paucity of our information, for example, dating a particularly revealing inscription to the fifth or third century BCE, or as originating from Athens or Asia Minor, could have serious scholarly ramifications. 1

Traditionally, scholars have used their own experience and specialized training to supply these missing contextual data [Emmanuel et al. 2021]. Recently, however, there has been increasing interest in using machine learning techniques to supplement, or even replace, subjective and idiosyncratic (although sometimes brilliant) evaluations. For example, Niculae et al. [Niculae et al. 2014] used machine learning techniques to date a corpus of older texts. 2

In 2022, Assael et al. [Assael et al. 2022] published their research into developing a machine learning platform that would aid the automated reconstruction and adding of missing contextual information to ancient Greek inscriptions. This platform, which they call Ithaca, is based on a deep neural network model. They demonstrate that such a technique greatly enhances scholarly expertise, although it cannot substitute for it. 3

As impressive as Ithaca is, deep neural network techniques presently have limited applicability to other digital humanities projects. One inherent problem with using them is that they are “black box” models; their processes remain opaque. Furthermore, they need both technical expertise to implement and a large sample size to train the algorithm [LeCun, Benigo, and Hinton 2015]. Datasets from antiquity, particularly those that exist in high-quality structured form, are rarely large enough to make this approach suitable. 4

In this paper, we explore the utility of other machine learning algorithms for predicting values in incomplete datasets. We have determined that a random forest model has the most potential to predict these values, particularly in smaller datasets with several categorical variables. While our own work was based on one dataset, “Inscriptions of Israel/Palestine,” we believe that our results are applicable to other datasets as well.

5

2. Methods

2.1 Inscription dataset

The “Inscriptions of Israel/Palestine” (IIP) dataset is an online database which seeks to make all of the previously published inscriptions of Israel/Palestine from the Persian period through the Islamic conquest (ca. 500 BCE - 640 CE) freely accessible [Satlow 2022]. This database includes approximately 4,500 inscriptions, and they are written primarily in Hebrew, Aramaic, Greek and Latin, by Jews, Christians, Greeks, and Romans. Some of the examples include imperial declarations on monumental architecture, notices of donations in synagogues and humble names scratched on ossuaries [Satlow 2022]. Each inscription exists as a single XML file structured according to EpiDoc conventions [Elliott, Bodard, Cayless, et al. 2006].

6

2.2 Variable Explanation

We consider the following characteristics in the dataset:

7

- Terminus ante quem (the latest possible date)
- Terminus post quem (the earliest possible date)
- Text Genre
- Language
- Material
- Region
- Likely Religion

It is worth noting that language, material, and region are objectively determined in most cases. Dating, on the other hand, is often determined subjectively by scholars; relatively few contain dates or were found in carefully controlled archaeological excavations. Thus, we examine how machine learning models can accurately predict the date of inscriptions given the information of other variables in the dataset. All variables inside the dataset except date are categorical, as they are not quantifiable.

8

2.3 Data Preprocessing

The IIP dataset is converted into a single csv file through using the ElementTree XML API in Python programming. One of the features of the IIP dataset, like many others in the humanities, is that it contains a number of categorical variables, that is, different phrases that occur within a single XML element. For example, there are many different cities in the location element, and over fifty different text genres (e.g., funerary, dedicatory, label, prayer) are found within the appropriate element. The result of this is an imbalanced dataset. Imbalanced datasets occur when the proportion of minority class is significantly low compared with other classes in the dataset, and creating an effective machine learning algorithm with imbalanced datasets is a very difficult problem [Johnson and Khoshgoftaar 2019]. It is also important to make sure that all the possible categorical values are in the training dataset to create a good machine learning model. For example, if the training set does not include any inscriptions that has the city name “Jerusalem”, it would be difficult for the machine learning algorithm to use the Jerusalem information in the test dataset. Error messages can come out in many machine learning programs when they encounter some information that is not included in the training dataset. Splitting the dataset into training and test set is done randomly, so it is possible for the minority class to not appear in both training and test set if the number of observations from the minority class is small. To have enough observations in the minority class, we combine various unique terms and generate a dataset that is better suited for the machine learning algorithm.

9

We first fix spelling mistakes inside the dataset. Afterwards, we combine words that describe the same concept. For instance, “Golan Heights” and “Golan” can be grouped together as “Golan” and there is no need for the machine learning algorithm to consider these elements separately. There are also some phrases such as “dedicatory quotation” and “dedicatory verse”, where they describe different objects but can be grouped together as “dedicatory” to reduce the number of variables inside the dataset. However, we take a different approach with the “City Name” variable. There are 244 unique city names inside the dataset, and many of them only include a few inscriptions. Since the location of inscription is already indicated in the “Region” variable, we only consider “Jerusalem” and “Other Cities” to make the prediction easier to interpret. We also do not consider all variables in the dataset, such as condition of artifacts and relief style, as they have a lot of missing values.

10

We use a technique called one-hot encoding to convert the categorical features to numerical features. There are many machine learning algorithms that can only analyze numerical data, so analyzing categorical variables without one-hot encoding can cause some issues. In this technique, we create a binary column for each category, where we denote the output of the column to be 1 if the variable is present and 0 if the variable is not present. For example, we create a new variable called `Language_Greek` to describe if the inscription is written in Greek or not. The `Language_Greek` variable will be 1 if the inscription is written in Greek, but 0 otherwise.

11

These inscriptions are used to examine the performance of machine learning models to predict the time periods. The time period distribution of inscription is shown in Figure 1. The mean date is 109.68 CE with standard deviation 311.47. The large standard deviation implies that the dataset is suitable for conducting this analysis, as it has inscriptions from a wide range of time periods. After we preprocess the dataset, we select 650 inscriptions that actually contain a certain date. The overview of the steps that are taken in this research project is shown in Figure 2.

12

2.4 Machine Learning Techniques

We compare the performances of eleven machine learning models: linear regression, ridge regression, lasso regression, elastic net, decision tree, random forest, neural network, XGBoost, and support vector regression with linear, radial and polynomial kernel. We select these algorithms, as they require minimal hyperparameter tuning and do not require data transformation. Hyperparameters determine the overall behavior of the machine learning model, and they must be set appropriately by the user before conducting the analysis [Claesen and DeMoor 2015]. Hyperparameter tuning is often performed manually, but it is impractical when we have many hyperparameters, and technical expertise is required to correctly set the hyperparameters [Claesen et al. 2014]. To create a simple and reproducible machine learning prediction model, we try to select models that do not require fine parameter tuning.

13

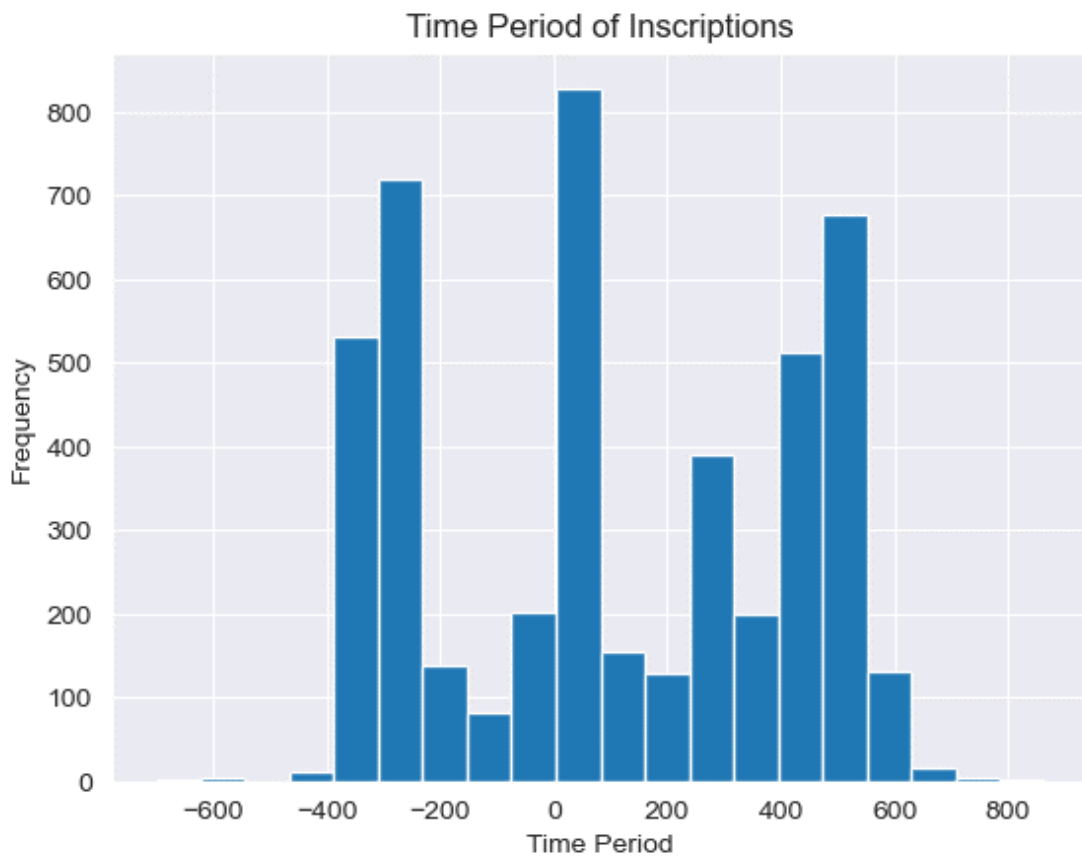


Figure 1. Time periods of inscriptions in the IIP dataset.

We will provide a brief overview of these techniques with some examples of previous studies in digital humanities. Readers who are interested in further details of machine learning techniques should consult Hastie et al. [Hastie et al. 2009].

2.4.1 Penalized Regression Techniques

Data Preprocessing

Convert an XML file to a CSV file through using the ElementTree XML API



Combine various unique terms in the dataset



Obtain the inscriptions where the precise date is known

Data Analysis

Chi-squared test to examine the association between the variables



Compare various machine learning techniques



Examine the random forest model through examining the variable importance plot

Figure 2. Workflow of the research project.

Ordinary least squares (OLS) regression is a commonly used statistical technique in regression problems. It assumes that there is a linear relationship between the predictor and response variable. We can directly observe the regression coefficients in OLS regression, so we can understand how the model is making predictions. There are, however, some disadvantages to using OLS. When the number of predictors become large, a small change in the training dataset can cause a large change in the prediction model produced by the OLS model [Hastie et al. 2009]. Thus, penalization techniques are often used to improve the predictability of OLS while retaining its linear model structure [Zou and Hastie 2005]. These methods impose a shrinkage penalty and bring the estimated coefficients closer to zero [Hastie et al. 2009]. We will be examining ridge, lasso, and elastic net, as they are commonly used penalization techniques.

15

Penalization techniques are frequently used in digital humanities research projects that contain datasets with many variables. For example, Finegold et al. [Finegold et al. 2016] used Poisson Graphical Lasso to reconstruct the historical social network in early modern Britain. They have imposed the penalization technique in statistical graph learning methods to find out the relationship between people's names inside the historical documents. Considering that the number of distinct names inside the historical documents is large, penalized regression went well for their analysis.

16

2.4.2 Support Vector Regression

Support Vector Regression (SVR) uses the same principle as Support Vector Machine (SVM), which is one of the most widely used supervised machine learning techniques [Drucker et al. 1996]. It is frequently used in digital humanities, including a study by Argamon et al. [Argamon et al. 2009], where they used SVM to classify author's gender from literary texts. The SVM algorithm conducts regression based on kernel functions, which converts the lower dimensional data into a higher dimensional feature space.

17

We consider the performances of three kernels, linear, radial and polynomial kernel, as they are commonly used kernels. The detailed information about the SVR mechanism can be found in [Drucker et al. 1996].

18

2.4.3 Neural Network

Deep learning algorithms make predictions based on a neural network structure, which is inspired by the human nervous system [Goodfellow, Benigo, and Courville 2016]. It has been used in multiple algorithms in digital humanities studies, including a study by Assael et al. [Assael et al. 2022] where they had implemented a deep learning algorithm to predict contextual information based on the textual information in ancient Greek inscriptions. To examine the performance of deep learning technique, we fit a single-hidden-layer neural network, as it has been shown that low complexity deep learning models perform better when the sample size is small [Brigato and Iocchi 2021].

19

2.4.4 Tree Based Approach

We examine three different tree based machine learning techniques, decision tree, random forest and Extreme Gradient Boosting (XGBoost). Random forest and XGBoost are tree ensemble methods, and they are considered to be the recommended tools to analyze tabular datasets [Borisov et al. 2022]. Ensembles are methods that combine multiple machine learning techniques to create more powerful models, and tree ensemble methods are used extensively in various digital humanities research. For example, a recent project by Baledent et al. [Baledent, Hiebel, and Lejeune 2020] used decision trees and random forests to automatically date French documents with high predictability. Fragkiadakis et al. [Fragkiadakis, Nyst, and Putten 2021] compared the performances of various machine learning techniques to annotate video data with sign languages, and showed that the XGBoost was the optimal model to predict the begin and end frames of a sign sequence in a video.

20

Decision tree is the foundation of random forest and XGBoost model. It is considered to be one of the most interpretable machine learning methods for data analysis, as it can classify data based on a set of yes/no questions. However, decision trees can be very non-robust and a minor change in the training data can result in a large change in the final tree [Hastie et al. 2009].

21

XGBoost is a tree ensemble machine learning algorithm that uses gradient boosted decision trees. It has a tree learning algorithm that enables to learn from sparse data, and it can analyze data faster than other popular machine learning techniques [Chen and Guestrin 2016]. The gradient boosting algorithm generates one tree at a time based on the previous model's residuals, and then they are combined to make the final prediction. In our analysis, we generate 150 trees in the final model, where the maximum depth of each tree is three.

22

The random forest algorithm is another tree ensemble machine learning algorithm that generates hundreds of decision trees by using a random subset of predictors in the bootstrapped samples. Bootstrapping is a statistical technique that repeatedly draw samples from the data with replacement [Hastie et al. 2009]. Since the same element can appear multiple times in the new sample, this technique generates a large number of new datasets that are not exactly the same as the original model. The average of the decision trees generated from the bootstrapped samples is examined to make the final prediction.

23

Random forest can also be used to rank the predictor variables based on its ability to decrease the sum of squared errors when it is chosen to split the data [Breiman 2001]. This is an important aspect of random forest, as we can understand which variables are important in the regression model to predict the criterion variable. Due to these advantages, multiple research highlight that random forests have emerged as serious competitors to other machine learning models for predicting numerical and categorical variables [Belgiu and Drăguț 2016].

24

To implement random forests, we only need to specify the number of trees and the number of features in each split. In terms of the number of trees, it has been shown that implementing many trees will provide a stable result of variable importance [Liaw and Wiener 2002] and using more than the required number of trees does not harm the model [Breiman 2001]. Many studies use $p/3$ number of features in each split for regression problems, where p is the number of predictor variables [Liaw and Wiener 2002].

25

2.5 Metric

Metrics are used to quantify the accuracy of the machine learning model once we obtain the machine learning models. We will examine three commonly used metrics to evaluate the machine learning algorithm, root mean square error

26

(RMSE), mean absolute error (MAE), and R-squared. 10-fold cross-validation is performed to compare the performances of machine learning algorithms. In k-fold cross validation, we split the dataset into k smaller sets with equal number of elements and use k-1 sets to train the model, while the remaining set is used to evaluate the model [Hastie et al. 2009]. We repeat the above iteration thirty times and compute the mean value of the determined metrics in cross validation to determine the optimal machine learning model for predicting the date.

2.6 Programming

We use R version 4.1.3 to perform the data analysis and Python version 3.8.3 to obtain the XML dataset from the IIP database [Satlow 2022]. The dataset and the codes that we use to obtain the dataset are openly available to the public at <https://github.com/daikitag/Inscriptions-of-Israel-Palestine>.

27

3. Results

3.1 Variable Relationship

It is important to understand the relationship between the predictor variables in the dataset before conducting machine learning analysis, as we can understand the issues behind effectively analyzing the dataset. Considering that all predictor variables are categorical, we use Pearson's chi-squared test of independence to examine the association between the variables in the dataset. We have examined the association between:

28

1. Language and Location
2. Religion and Location
3. Religion and Language
4. Religion and Text Genre

The residual plots of the chi-squared test are shown in Figure 3. Results from chi-squared test indicate that there is a significant relationship between all the examined combinations ($p < 0.001$).

29

The results imply that the predictor variables are correlated with each other, which raises the problem of multicollinearity. Multicollinearity occurs when independent variables in the regression model are correlated with each other [Alin 2010]. One of the key assumptions of the linear regression model is that the predictor variables are uncorrelated. Thus, multicollinearity can undermine the statistical significance of an independent variable and can give inaccurate coefficient estimates when traditional statistical techniques are used [Allen 1997].

30

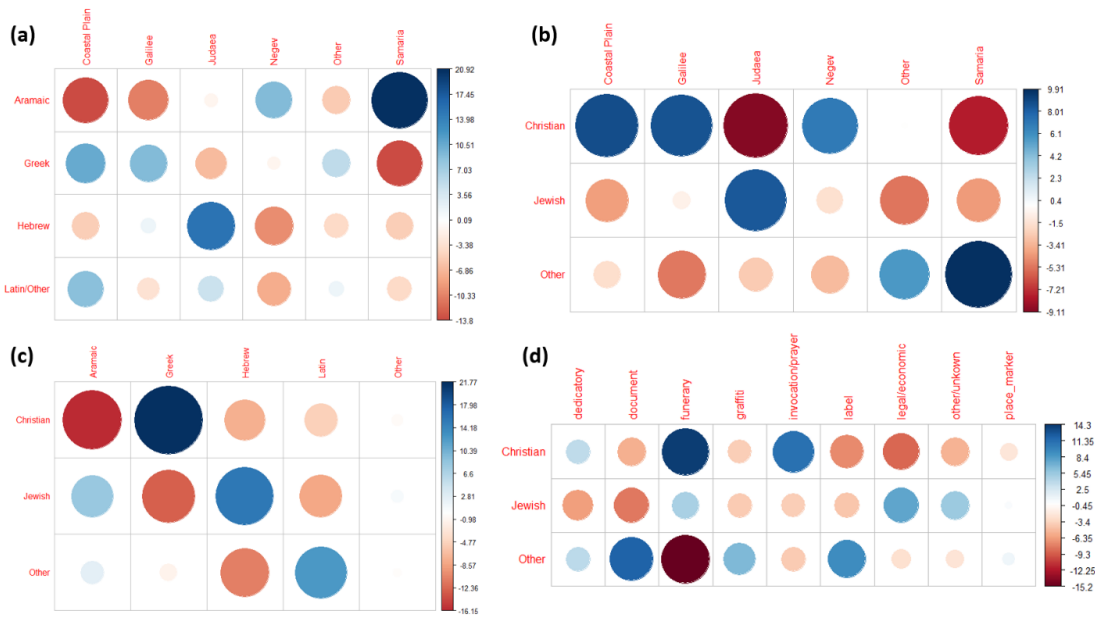


Figure 3. Residual plot of chi-squared analysis of the dataset. Red color indicates that two variables are negatively associated, and blue color indicates that two variables are positively associated. (a) Chi-squared test between language and location of inscriptions. (b) Chi-squared test between religion and location. (c) Chi-squared test between religion and language. (d) Chi-squared test between religion and text genre.

In contrast, due to recent advances in machine learning techniques, it has been reported that machine learning can better analyze data with multicollinearity than traditional statistical techniques [Chan et al. 2022]. The presence of multicollinearity suggests the usage of machine learning techniques to effectively analyze the dataset.

31

3.2 Machine Learning Model Comparison

A total of eleven regression models are compared: linear regression, ridge regression, lasso regression, elastic net, decision tree, random forest, neural network, SVR linear, SVR radial, SVR polynomial, and XGBoost. The optimal hyperparameters of the machine learning algorithms are determined through 10-fold cross-validation. The cross-validation procedure is repeated 3 times, and the machine learning model is tested by using a total of 30 different datasets, each of which are generated through cross-validation. The evaluation results are shown in Figure 4, where Figure 4 (a) shows the distribution of RMSE from 30 different datasets and Figure 4 (b) shows the mean values of MAE, RMSE and R-squared. MAE and RMSE measure the error of the machine learning model and R-squared is a goodness of fit measure. The random forest model has the lowest value for MAE and RMSE, and has the highest value for R-squared among all models that are examined. This implies that random forest model is the optimal model for predicting the date of inscriptions.

32

3.3 Random Forest Model

We describe the random forest model in detail, as it is the best model that is implemented in the previous section. We initially convert the number of trees that the random forest model generates from 100 to 1000 to determine the optimal number of trees that we put inside the algorithm, but we do not observe any significant differences. Thus, we select 500 number of trees, as it is the default number of trees in R's randomForest package [Liw and Wiener 2002].

33

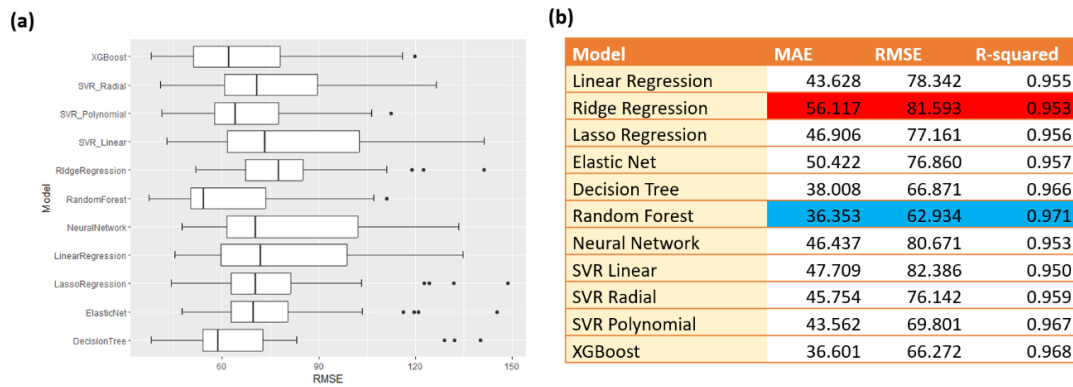


Figure 4. Evaluation of machine learning methods from 10-fold cross validation. (a) The distribution of RMSE from cross validation. (b) Mean values of MAE, RMSE and R-squared from cross validation. The best model for each metric is colored by blue and the worst model is colored by red.

Figure 5 (a) shows the variable importance plot of the random forest model. Variable importance is based upon the mean increase of mean squared error as a result of permuting a given variable [Liau and Wiener 2002]. The plot suggests that the material of inscription is the most important variable in the prediction model. The prediction plot is shown in Figure 5 (b). The machine learning model is trained based on the training dataset, which includes 70% of the randomly chosen inscriptions in the data. The prediction plot is created by using the test dataset, which is not used to train the machine learning model.

34

4. Discussions

4.1 Categorical Analysis

Every region and community in the Mediterranean in antiquity had its own epigraphic characteristics. The statistical analysis reveals some features of different communities within Judea/Roman Palestine. From the residual plot in Figure 3 (a), we can infer that inscription in Judea have a higher probability of being written in Hebrew and inscriptions in Negev have a higher probability of being written in Aramaic. There is also a very strong positive association between Aramaic and Samaria. However, there is a lower probability of Aramaic inscriptions found in the Coastal Plain.

35

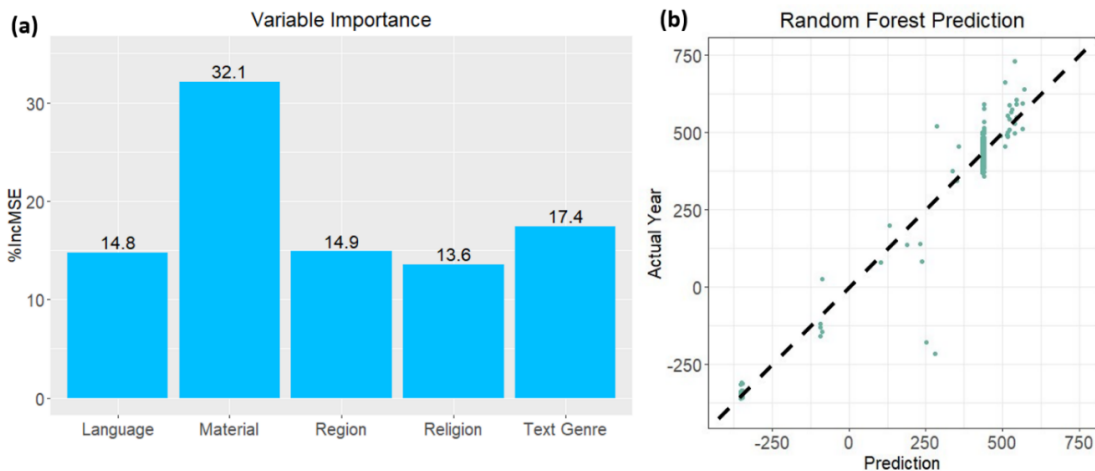


Figure 5. (a) Variable importance plot of the random forest model. (b) The relationship between prediction and actual value of the test dataset. The dashed line represents the location where the prediction and the actual value are the same.

When we examine the residual plot in Figure 3 (b), there is a higher possibility of discovering Christian inscriptions in Coastal Plain, Galilee, and Negev. There is a higher possibility of discovering Jewish inscriptions in Judea and Galilee.

36

However, the probability of finding Christian and Jewish inscriptions in Samaria is lower than other regions, and there are many inscriptions from other religions. These results are consistent with what we would expect from other historical sources.

The residual plot of language and religion is shown in Figure 3 (c). Christian inscriptions have a strong positive association with Greek, but negative association with other languages, specifically Aramaic. Inscriptions written in Aramaic and Hebrew are more likely to be Jewish inscriptions. 37

The relationship between text genre and religion is shown in Figure 3 (d). The plot implies that Christian inscriptions tend to be funerary or invocation related compared with other religions, but the probability of Christian inscription being document or legal/economic is lower. It seems that Christian inscriptions tend to be more religious and less administrative. 38

4.2 Machine Learning Model

We are able to conclude that the random forest model is the optimal machine learning model for predicting time periods of inscriptions. This is consistent with previous research, as it has been reported that tree-ensemble algorithms like random forests are better to analyze tabular data [Shwartz-Ziv and Armon 2022], which is the data type that we use in our project. If we examine the metric values in Figure 4 (b), we see that random forest, XGBoost and decision tree perform better than other models that we have examined. This highlights the importance of using tree based machine learning algorithms to analyze tabular dataset. 39

Our study also shows that linear models do not perform well compared with other methods in analyzing tabular datasets which consists of only categorical variables. This might be due to the nonlinear interactions between the variables in the dataset. In terms of SVM, it is important to select the appropriate kernel for each dataset. In our example, we see that SVR with linear kernel performs the worst out of all three kernels that we have examined. However, there are many kinds of kernels in SVM, and it would be a challenging problem to select the optimal kernel for the dataset. The results also suggest that the predictability of neural network is not high compared with tree-based algorithms when we analyze tabular datasets. Many digital humanities datasets are tabular data and they are rarely large enough to effectively train the deep learning algorithm. Our results are consistent with the previous study by Shwartz-Ziv & Armon [Shwartz-Ziv and Armon 2022], where they also showed that tree ensemble methods are better than deep learning techniques to analyze tabular data. 40

In contrast, a random forest model can easily be implemented by specifying two hyperparameters of the model, and it is possible for a random forest model to capture the nonlinear interactions inside the dataset. This is a major advantage of random forests, as most machine learning models require fine tuning of hyperparameters [Hastie et al. 2009]. 41

In spite of advances in machine learning techniques, it is still necessary to have epigraphers to analyze inscriptions. According to the variable importance plot of the random forest model (Figure 5 (b)), material is the most important variable in making predictions, but we cannot ignore the effects of other variables, including religion and text genre. These variables are subjective, and necessitates the importance of having humans to classify the inscriptions as well. Even in the research project conducted by Assael et al. [Assael et al. 2022], they showed that accuracy was the best when the deep learning algorithm was paired up with historians. Machine learning algorithms are still not perfect, so it would be important for us to incorporate knowledge from both human scholars and computers to analyze the dataset effectively. 42

5. Conclusions

We show how machine learning techniques can be used to make predictions based on tabular dataset that is comprised of categorical variables. It is uncommon for humanities data to include all elements of a dataset. This could be due to the damage of artifacts over time and many texts being often only available in fragments. Instead of only using one element of the dataset to make predictions, it would be important for us to incorporate other elements in the dataset to effectively date the artifacts. As a next step, we plan to integrate the deep learning framework to the machine learning 43

model that we have created, so that we can incorporate both textual data and tabular data of inscriptions in the prediction model to achieve better accuracy.

The results of our work indicate that computers can successfully be taught to predict missing characteristics of historical artifacts. The widespread use of machine learning techniques offers exciting prospects in epigraphy and related fields. Even if the dataset is not large, we provide an example in which machine learning techniques can effectively be used to make predictions. In addition to the inscription dataset, our research shows that the machine learning model could be used to analyze other digital humanities dataset which includes a wide range of categorical variables.

44

Acknowledgements

We acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010. We thank Brown University's Center for Digital Scholarship for providing the valuable dataset for our research. We would also like to thank the anonymous reviewers, as their suggestions and comments have significantly improved the content and presentation of this paper.

45

Works Cited

- Alin 2010** Alin, A. (2010) "Multicollinearity." *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), pp. 370–374.
- Allen 1997** Allen, M. P. (1997) "The problem of multicollinearity". In Allen, M.P. *Understanding Regression Analysis*. Berlin: Springer, pp. 176-180.
- Argamon et al. 2009** Argamon, S., Goulain, J.-B., Horton, R., and Olsen, M. (2009) "Vive la différence! Text mining gender difference in french literature." *Digital Humanities Quarterly*, 3(2).
- Assael et al. 2022** Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., and de Freitas, N. (2022) *Restoring and attributing ancient texts using deep neural networks*. *Nature*, 603(7900), pp. 280–283.
- Baledent, Hiebel, and Lejeune 2020** Baledent, A., Hiebel, N., and Lejeune, G. (2020) "Dating ancient texts: An approach for noisy French documents." *Language Resources and Evaluation Conference (LREC) 2020*.
- Belgiu and Drăguț 2016** Belgiu, M., and Drăguț, L. (2016) Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31.
- Borisov et al. 2022** Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022) "Deep Neural Networks and Tabular Data: A Survey." *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21. <https://doi.org/10.1109/TNNLS.2022.3229161>
- Breiman 2001** Breiman, L. (2001). "Random forests." *Machine Learning*, 45(1), pp. 5–32.
- Brigato and Iocchi 2021** Brigato, L., and Iocchi, L. (2021) "A Close Look at Deep Learning with Small Data." *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2490–2497. <https://doi.org/10.1109/ICPR48806.2021.9412492>
- Chan et al. 2022** Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., and Chen, Y. L. (2022) "Mitigating the multicollinearity problem and its machine learning approach: a review". *Mathematics*, 10(8), 1283.
- Chen and Guestrin 2016** Chen, T., and Guestrin, C. (2016) "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Claesen and DeMoor 2015** Claesen, M., and De Moor, B. (2015) *Hyperparameter Search in Machine Learning* (arXiv:1502.02127). arXiv. <https://doi.org/10.48550/arXiv.1502.02127>
- Claesen et al. 2014** Claesen, M., Simm, J., Popovic, D., Moreau, Y., and De Moor, B. (2014) *Easy Hyperparameter Search Using Optunity* (arXiv:1412.1114). arXiv. <https://doi.org/10.48550/arXiv.1412.1114>
- Drucker et al. 1996** Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996) "Support vector regression machines." *Advances in Neural Information Processing Systems*, 9.

- Elliott, Bodard, Cayless, et al. 2006** Elliott, Tom, Bodard, Gabriel, and Cayless, Hugh et al. (2006, 2022) *EpiDoc: Epigraphic Documents in TEI XML*. <https://epidoc.stoa.org/>
- Emmanuel et al. 2021** Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O. (2021) "A survey on missing data in machine learning." *Journal of Big Data*, 8(1), pp. 1–37.
- Finegold et al. 2016** Finegold, M., Otis, J., Shalizi, C., Shore, D., Wang, L., and Warren, C. (2016) "Six degrees of Francis Bacon: A statistical method for reconstructing large historical social networks." *Digital Humanities Quarterly*, 10(3).
- Fragkiadakis, Nyst, and Putten 2021** Fragkiadakis, M., Nyst, V., and Putten, P. van der. (2021) "Towards a User-Friendly Tool for Automated Sign Annotation: Identification and Annotation of Time Slots, Number of Hands, and Handshape." *Digital Humanities Quarterly*, 15(1).
- Goodfellow, Benigo, and Courville 2016** Goodfellow, I., Bengio, Y., and Courville, A. (2016) *Deep Learning*. MIT press.
- Hastie et al. 2009** Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2). Springer.
- Johnson and Khoshgoftaar 2019** Johnson, J. M., and Khoshgoftaar, T. M. (2019) "Survey on deep learning with class imbalance." *Journal of Big Data*, 6(1), 27. <https://doi.org/10.1186/s40537-019-0192-5>
- LeCun, Benigo, and Hinton 2015** LeCun, Y., Bengio, Y., and Hinton, G. (2015) "Deep learning." *Nature*, 521 (7553), pp. 436–444.
- Liaw and Wiener 2002** Liaw, A., and Wiener, M. (2002) "Classification and regression by randomForest." *R News*, 2(3), pp. 18–22.
- Niculae et al. 2014** Niculae, V., Zampieri, M., Dinu, L., and Ciobanu, A. M. (2014) "Temporal Text Ranking and Automatic Dating of Texts." *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers*, pp. 17–21. <https://doi.org/10.3115/v1/E14-4004>
- Satlow 2022** Satlow, M. L. (2022) "Inscriptions of Israel/Palestine." *Jewish Studies Quarterly (JSQ)*, 29(4), pp. 349–369. <https://doi.org/10.1628/jsq-2022-0021>
- Shwartz-Ziv and Armon 2022** Shwartz-Ziv, R., and Armon, A. (2022) "Tabular data: Deep learning is not all you need." *Information Fusion*, 81, pp. 84–90.
- Zhitomirsky-Geffet et al. 2020** Zhitomirsky-Geffet, Maayan, Gila Prebor, and Isaac Miller. (2020) "Ontology-based analysis of the large collection of historical Hebrew manuscripts." *Digital Scholarship in the Humanities*, 35(3), pp. 688–719. <https://doi.org/10.1093/llc/fqz058>
- Zou and Hastie 2005** Zou, H., & Hastie, T. (2005). "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), pp. 301–320.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.