The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 22, Pages 237-246

ICBASET 2023: International Conference on Basic Sciences, Engineering and Technology

Variable Selection with Machine Learning in the Legalization Process for Traffic Insurance

Vedat GUNES
Anadolu Insurance

Serkan KIRCA
Anadolu Insurance

Hasan Ersan YAGCI
Muğla Sıtkı Koçman University

Nida Gokce NARIN
Muğla Sıtkı Koçman University

Abstract: In the insurance sector, the insured notifies the insurance company of which he is the customer as soon as the damage occurs. Upon this notice, a claim file is opened to the insured, and the damage file number is assigned. The claim file contains information about the product insured by the insured and the damage. This information is kept in tables in the databases of Anadolu Insurance. In the event of damage, the insured's claim can be accepted. The entire damage amount can be paid if the damage amount is partially accepted, with the examination to be carried out by the insurance company; if the damage amount is partially accepted, a part of it is paid, or the claim is rejected. The damage amount is not paid at all. When the Insured receives partial payment or the claim file is rejected, they can sue the insurance company to claim the damage amount. The litigation process is a long and bad experience for the insured. For the insurance company, in addition to customer dissatisfaction, it causes extra costs such as court, lawyer, etc. costs. The problem studied in this work is aimed to determine which variables are essential for a possible legalization process in case of partial acceptance or rejection of the claim file by using the variables in the relevant claim file by machine learning and statistical methods. While making this determination, lasso regression, information gain, chi-square test, fisher's score, Recursive Feature Elimination (RFE) with Random Forest Machine Learning algorithm, Univariate Feature Selection with bivariate statistical tests or univariate statistics like chi-square test and feature importance of Random Forest Machine Learning algorithm. Variable selection was made by using correlation coefficient and backward feature elimination methods. Variable p_value was also evaluated.

Keywords: Insurance, Legalization, Anadolu insurance, Variable selection, Machine learning, statistics

Introduction

In the insurance sector, the process of preparing a claim can be a complex and challenging experience for both the insured and the insurance company. When an insured suffers damage to their insured property, they must notify the insurance company immediately to open a claim file. The file contains information about the product insured and the damage suffered. Based on the examination of the claim file, the insurance company can either accept the claim and pay the entire damage amount or partially accept the claim and pay a portion of the amount. If the claim is rejected, the insured can sue the insurance company to claim the damage amount, leading to a lengthy and costly litigation process.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

The purpose of this study is to determine which variables in the claim file are essential in predicting the likelihood of a potential legalization process in case of partial acceptance or rejection of the claim file. Machine learning and statistical methods such as lasso regression, information gain, chi-square test, fisher's score, Recursive Feature Elimination (RFE) with Random Forest Machine Learning algorithm, Univariate Feature Selection with bivariate statistical tests or univariate statistics like chi-square test and feature importance of Random Forest Machine Learning algorithm were used to select relevant variables. The variable selection process was performed using correlation coefficient and backward feature elimination methods.

The results of this study will provide valuable insights into which variables play a significant role in determining the outcome of a claim file and can aid insurance companies in making informed decisions. By identifying relevant variables, insurance companies can reduce the likelihood of a litigation process and minimize costs associated with legal fees and customer dissatisfaction.

Claim Process

The damage process starts with opening the claim file as soon as the damage and alkali notification is received and continues until the opened damage file is closed. In this process, it is aimed to ensure customer satisfaction and prevent possible costs by tracking the cost of damage. In the damage process, insurance companies have a great responsibility. The main of these responsibilities is to repair,

- Fast
- Flawless
- Complete

to ensure that it is done. In the study conducted here, the analysis and inference of the variables that will affect the legalization of the traffic damages belonging to the traffic branch in the damage process are defined as auto damage.

The legalization process can sue the Insurance company to claim the damage amount when the insured receives partial payment, or the claims file is denied. The litigation process is a long and bad experience for the insured. For the insurance company, in addition to customer dissatisfaction, it causes extra costs such as court, lawyer, etc. costs. In addition, the insured can be sued not only by the insured but also by the insurance company. For this, detecting irregularities that occur in accidents can be given as an example.

The following processes in the legalization process:

- First of all, there is a valid (valid) policy belonging to the customer.
- Occurrence of damage (i.e. hitting a tree or collision with vehicles)
- Keeping the damage reports and informing the insurance company
- Showing the vehicle to the service
- The insurance company appoints a loss adjuster, and the expert sees the intermediary at the service and starts the repair process.
- By writing the loss adjuster's report, the reflection of the loss adjuster's report to the insurance company's system
- Completing the repair process and informing the insured

sequentially, the damage process is completed.

Machine Learning Methods

The problem addressed is whether the customer of the relevant claim file will take legal action when the file is rejected or whether payment is made partially by using the variables in the report as soon as the loss adjuster's reports are reflected in the Anadolu Sigorta system. In this study, this process will be called legalization. To solve this problem, the classification method, one of the machine learning methods, was determined as the most suitable method for the problem (Uğurlu et al.,).

The subject of the study corresponds to the Supervised Learning – Classification section.

The high success score of the classification accuracy is highly correlated with the proper selection of the variables. The variable assignment is the most critical process affecting the model performance and accuracy. It should be known that when the data set is prepared, not all of the independent variables (attributes) in the data set are meaningful for the model (Kaya & Köymen, 2008). For this reason, the selection of variables should be handled before creating the model. The meaningful variables for the model should be selected, and modeling studies should be carried out specific to these variables.

Method

Preparing the Data Set Suitable for the Problem

A study-specific datamart (Li & Liu, 2017) was created using the data created in the claim process and stored in the databases belonging to Anadolu Insurance database. In the datamart developed, there are 157 variables; 156 are features, and 1 is the target variable specific to the problem. The superset of the attributes created in the data model is 810. The target variable preaches the claim file legalized (Yes / No). It was created according to the binary classification method. All the data preparation steps were developed by using IBM's Infosphere Datastage. The flow in the Datastage platform is shown in Figure 2.

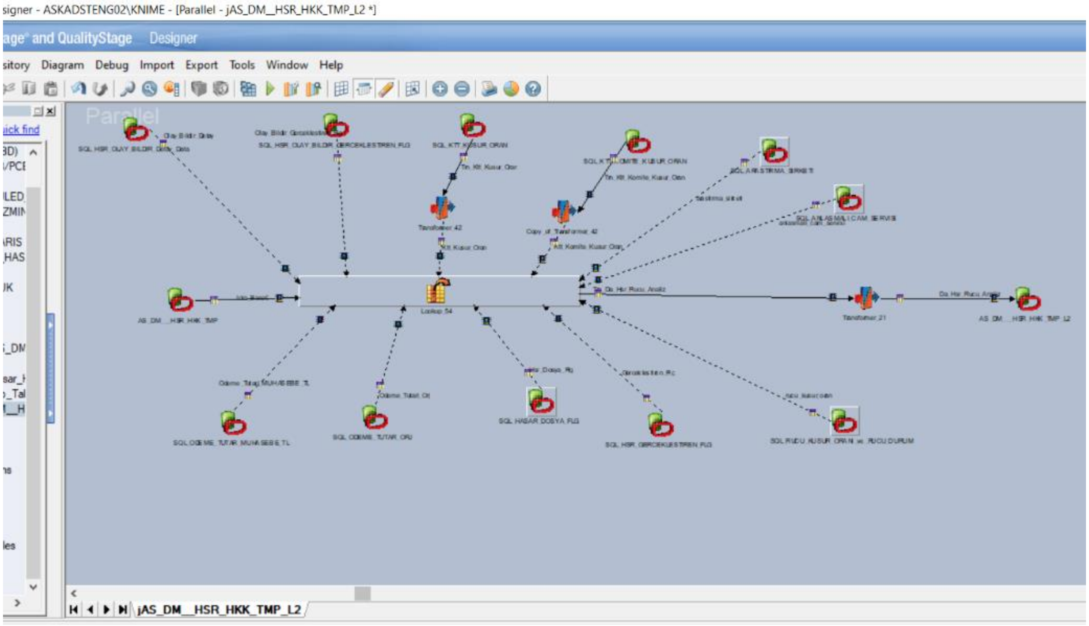


Figure 1. IBM's infosphere datastage platform

Feature Selection Methods

When the literature is examined, statistical and variable selection methods are based on machine learning algorithms (Zhang et al., 2021). Some of these algorithms are;

- Lasso Regression Feature Selection,
- Information Gain,
- Chi-Square Test,
- Fisher's Score,
- Recursive Feature Elimination (RFE),
- Univariate Feature Selection,
- Feature importance of Random Forest ML Algorithm,
- Correlation Coefficient,
- Backward Feature Elimination

methods and more are available in the literature. Let us briefly explain the above variable selection methods.

Lasso Regresyon Feature Selection

It consists of adding a penalty to different parameters of the machine learning model to reduce the freedom of the model, that is, to prevent overfitting (Kozak et al., 2020). In the linear model arrangement, the penalty is applied over the coefficients that hit each estimator. Lasso or L1 can reduce some coefficients to zero of the different regularization types. Therefore, features reduced to zero are removed from the data set before model creation.

Information Gain

Information Gain calculates the reduction in entropy from the transformation of a dataset (İlhan & Sarı, 2016).It can be used for feature selection by evaluating the information gain method outputs for each variable in the context of the target variable.

Chi-Square Test

The chi-square test is used for categorical features in a data set (Santos et al., 2012). We calculate Chi-square between each feature and target. Then select the desired number of features with the best Chi-square scores. The following conditions must be met to correctly apply chi-square to test the relationship between the various features in the dataset and the target variable. The variables should be categorical and sampled independently, and the expected frequency of the values should be higher.

Fisher’s Score

The Fisher score is one of the most widely used supervised feature selection methods. The algorithm we will use returns the order of the variables in descending order according to the fisher's score. Then the variables are selected according to the problem (Sharma & Goyal, 2020).

Recursive Feature Elimination (RFE)

Recursive Feature Elimination is a wrapper-type feature selection algorithm. It means that a different machine learning algorithm is given and used at the method's core, wrapped by RFE and used to help select features. RFE, contrasts filter-based feature selections, which score each feature and choose the elements with the highest (or smallest) score. Technically, RFE is a wrapper-style feature selection algorithm that also uses filter-based feature selection internally (Garcia et al., 2017). RFE works by searching for a subset of features, starting with all features in the training dataset and successfully removing features until the desired number remains.

The final feature set is achieved by fitting the specific machine learning algorithm used at the model's core, ranking features by importance, discarding the least important features, and refitting the model. This process is repeated until a certain number of attributes remain.

Univariate Feature Selection

Univariate feature selection runs by selecting the best features based on univariate statistical tests. We compare each feature with the target variable to see if there is a statistically significant relationship between them. It is also called analysis of variance (ANOVA) (Gregorutti & Michel, 2017). We ignore other features when examining the relationship between a feature and the target variable. Because of the application steps, it is called 'univariate.'

Feature importance of Random Forest ML Algorithm

Tree-based strategies used by random forests are naturally ranked by how well they improve the purity of the node, or in other words, the reduction in impurity (Gini impurity) across all trees (Kirisci, 2022). The nodes with the most significant reduction in impurity occur at the beginning of the trees, and the notes with the most negligible reduction in impurity appear at the end of the trees. Thus, pruning trees below a particular node can generate a subset of the most important features.

Correlation Coefficient

Correlation is a degree of the linear relationship between two or more variables (Aker, 2022). Through correlation, it can predict one variable from another. The purpose of using correlation for feature selection is that relevant variables correlate highly with the target variable. Also, the variables should be correlated with the target variable but uncorrelated among themselves. If two variables are correlated, one is estimated from the other. Therefore, if two properties are associated, the model only needs one, as the latter does not add additional information.

Backward Feature Elimination

This method works oppositely to the Advanced Feature Selection method (Akar, 2021). Here, we start with all the available features and build a model. Next, we take the variable from the model that gives the best evaluation measure value. This process is continued until the predetermined criterion is met [15].

Results and Discussion

Some specified variable selection algorithms have been selected and explained under a heading. The following variable selection algorithms were used in this study.

- Feature Importance of the Random Forest ML Algorithm
- Lasso Regression
- Information Gain
- Correlation Coefficient

Feature Importance of Random Forest ML Algorithm

When we apply the random forest feature selection algorithm to the data set, the results in figure 2 and table 1 are obtained.

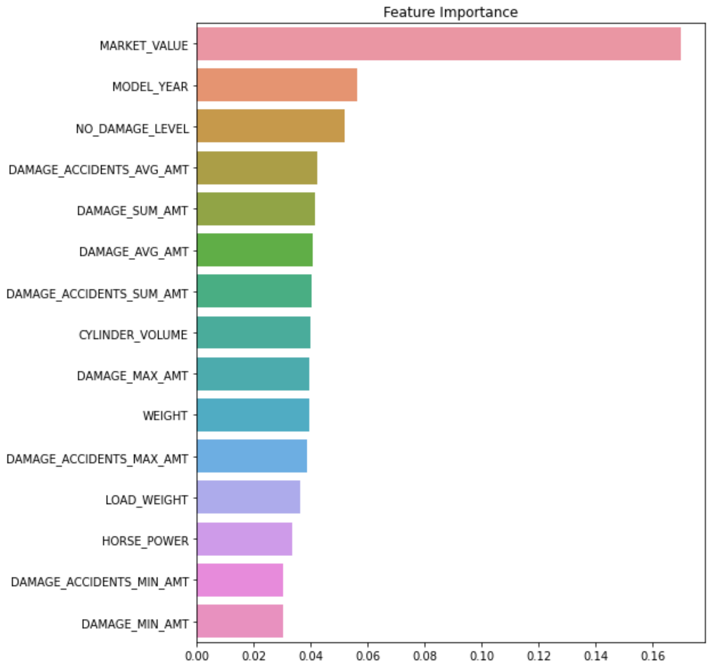


Figure 2. Random forest feature selection algorithm figural results

Table 1. Random forest feature selection algorithm values

Variables	Feature Importances
MARKET_VALUE	0.167
MODEL_YEAR	0.057
NO_DAMAGE_LEVEL	0.051
DAMAGE_ACCIDENTS_AVG_AMT	0.043
DAMAGE_SUM_AMT	0.042
DAMAGE_ACCIDENTS_SUM_AMT	0.041
CYLINDER_VOLUME	0.039
DAMAGE_MAX_AMT	0.039
DAMAGE_ACCIDENTS_MAX_AMT	0.039
WEIGHT	0.039
LOAD_WEIGHT	0.036
HORSE_POWER	0.034
DAMAGE_MIN_AMT	0.030
DAMAGE_ACCIDENTS_MIN_AMT	0.030

When Table 1 is examined, the essential variables are listed from the largest to the smallest at the variable importance level. The most important variable is the vehicle's market value, as expected. Afterward, the model year continues as the undamaged tier.

Lasso Regression

It is a regression technique in which variable selection and regularization co-occur. It is widely applied in large datasets due to its efficiency and speed. The results obtained are shown in Figure 3.

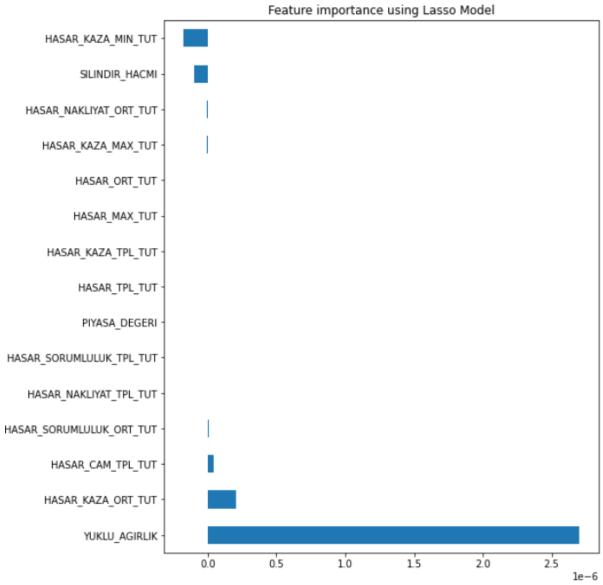


Figure 3. Lasso regression results

Figure 3 shows the variables with positive and negative effects from variable significance levels. While the minimum amount of the Caspian accident has a negative impact, the loaded weight has a positive significance, but the variable of average damage has no significant contribution.

Information Gain

It is a regression technique in which both variable selection and regularization take place at the same time. It is widely applied in large datasets due to its efficiency and speed. Information Gain calculates the reduction in entropy from the transformation of a dataset. It can be used for feature selection by evaluating the information gain results of each variable in the context of the target variable. The results obtained are shown in Figure 4.

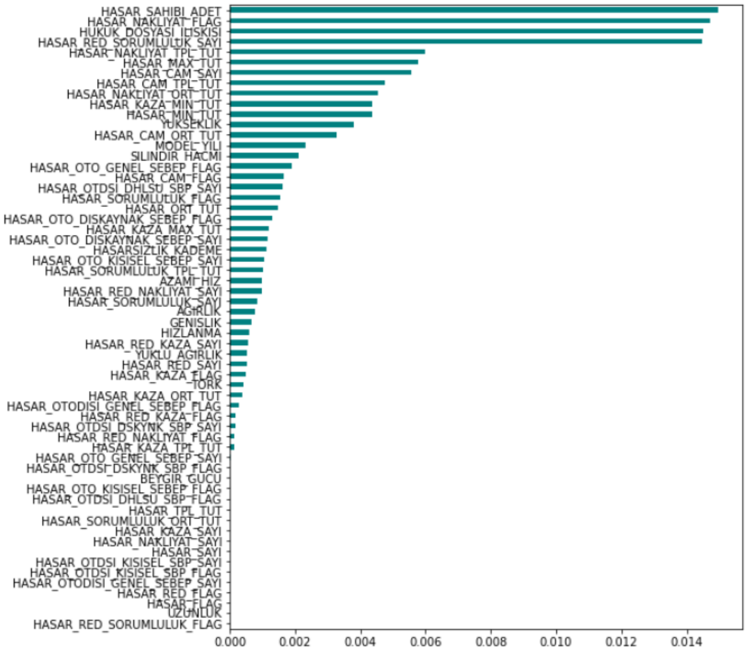
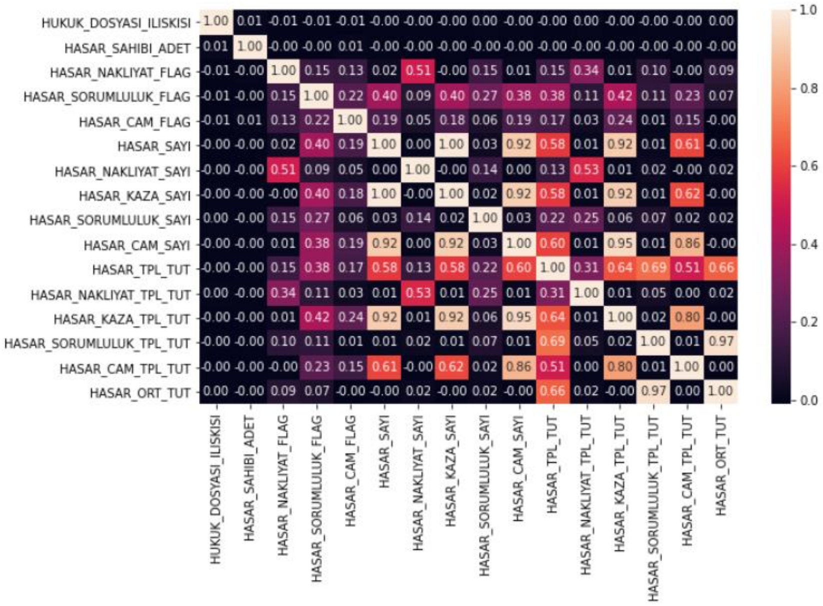


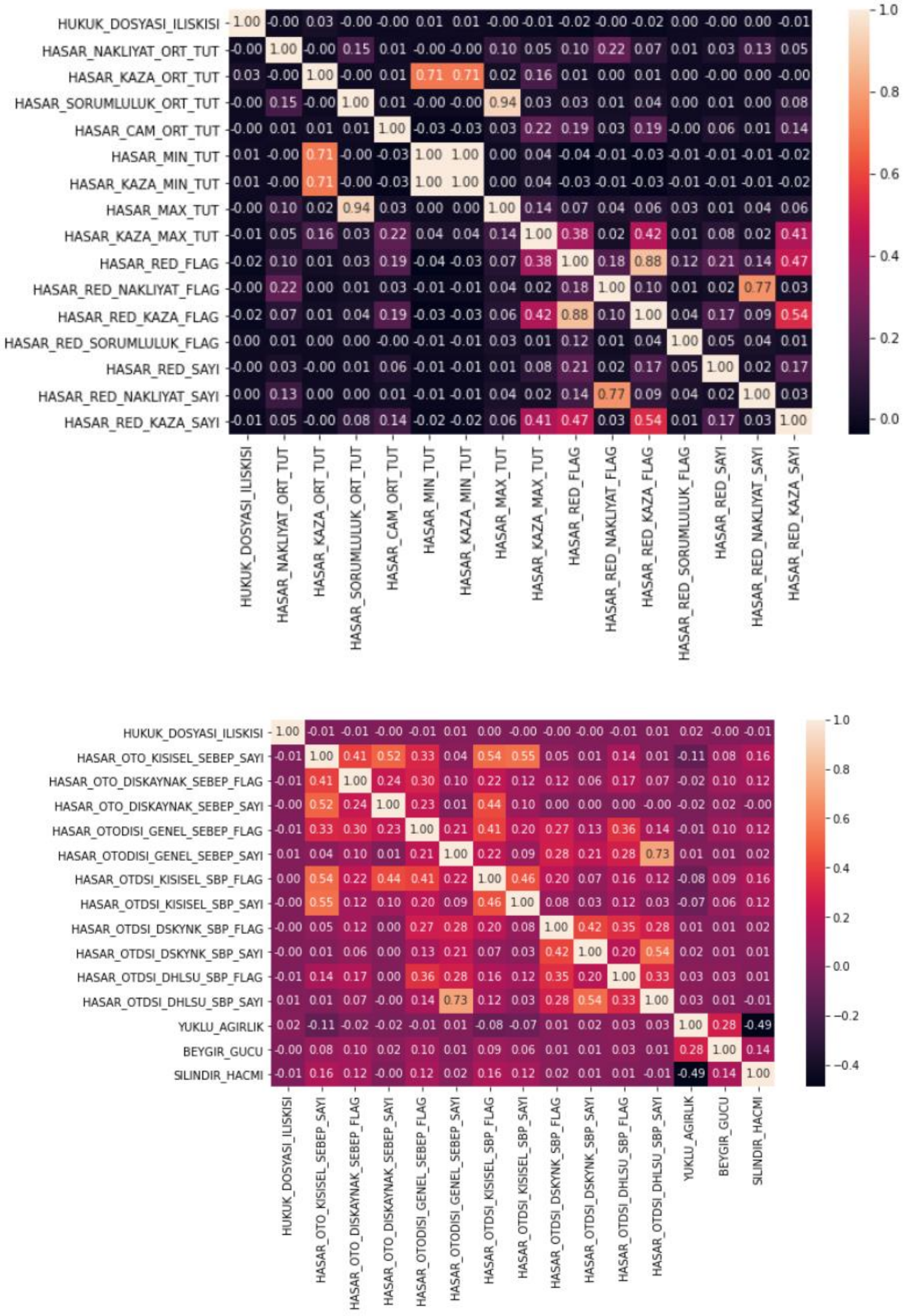
Figure 4. Information gain results

When Figure 4 is examined, the variables whose significance level is important are listed from the largest to the smallest. While the number of claimants is a vital variable, other variables are ranked according to the level of importance.

Correlation Coefficient

Correlation is a measure of the linear relationship between two or more variables. Through correlation, we can predict one variable from another. The rationale behind using correlation for feature selection is that suitable variables are highly correlated with the target. Also, the variables should be correlated to the target but uncorrelated among themselves. The results are shown in Figure 5.





	HUKUK_DOSYASI_ILISKISI	HASAR_NAKLIYAT_ORT_TUT	HASAR_KAZA_ORT_TUT	HASAR_SORUMLULUK_ORT_TUT	HASAR_CAM_ORT_TUT	HASAR_MIN_TUT	HASAR_KAZA_MIN_TUT	HASAR_MAX_TUT	HASAR_KAZA_MAX_TUT	HASAR_RED_FLAG	HASAR_RED_NAKLIYAT_FLAG	HASAR_RED_KAZA_FLAG	HASAR_RED_SORUMLULUK_FLAG	HASAR_RED_SAYI	HASAR_RED_NAKLIYAT_SAYI	HASAR_RED_KAZA_SAYI
HUKUK_DOSYASI_ILISKISI	1.00	-0.00	0.03	-0.00	-0.00	0.01	0.01	-0.00	-0.01	-0.02	-0.00	-0.02	0.00	-0.00	0.00	-0.01
HASAR_NAKLIYAT_ORT_TUT	-0.00	1.00	-0.00	0.15	0.01	-0.00	-0.00	0.10	0.05	0.10	0.22	0.07	0.01	0.03	0.13	0.05
HASAR_KAZA_ORT_TUT	0.03	-0.00	1.00	-0.00	0.01	0.71	0.71	0.02	0.16	0.01	0.00	0.01	0.00	-0.00	0.00	-0.00
HASAR_SORUMLULUK_ORT_TUT	-0.00	0.15	-0.00	1.00	0.01	-0.00	-0.00	0.94	0.03	0.03	0.01	0.04	0.00	0.01	0.00	0.08
HASAR_CAM_ORT_TUT	-0.00	0.01	0.01	0.01	1.00	-0.03	-0.03	0.03	0.22	0.19	0.03	0.19	-0.00	0.06	0.01	0.14
HASAR_MIN_TUT	0.01	-0.00	0.71	-0.00	-0.03	1.00	1.00	0.00	0.04	-0.04	-0.01	-0.03	-0.01	-0.01	-0.01	-0.02
HASAR_KAZA_MIN_TUT	0.01	-0.00	0.71	-0.00	-0.03	1.00	1.00	0.00	0.04	-0.03	-0.01	-0.03	-0.01	-0.01	-0.01	-0.02
HASAR_MAX_TUT	-0.00	0.10	0.02	0.94	0.03	0.00	0.00	1.00	0.14	0.07	0.04	0.06	0.03	0.01	0.04	0.06
HASAR_KAZA_MAX_TUT	-0.01	0.05	0.16	0.03	0.22	0.04	0.04	0.14	1.00	0.38	0.02	0.42	0.01	0.08	0.02	0.41
HASAR_RED_FLAG	-0.02	0.10	0.01	0.03	0.19	-0.04	-0.03	0.07	0.38	1.00	0.18	0.88	0.12	0.21	0.14	0.47
HASAR_RED_NAKLIYAT_FLAG	-0.00	0.22	0.00	0.01	0.03	-0.01	-0.01	0.04	0.02	0.18	1.00	0.10	0.01	0.02	0.77	0.03
HASAR_RED_KAZA_FLAG	-0.02	0.07	0.01	0.04	0.19	-0.03	-0.03	0.06	0.42	0.88	0.10	1.00	0.04	0.17	0.09	0.54
HASAR_RED_SORUMLULUK_FLAG	0.00	0.01	0.00	0.00	-0.00	-0.01	-0.01	0.03	0.01	0.12	0.01	0.04	1.00	0.05	0.04	0.01
HASAR_RED_SAYI	-0.00	0.03	-0.00	0.01	0.06	-0.01	-0.01	0.01	0.08	0.21	0.02	0.17	0.05	1.00	0.02	0.17
HASAR_RED_NAKLIYAT_SAYI	0.00	0.13	0.00	0.00	0.01	-0.01	-0.01	0.04	0.02	0.14	0.77	0.09	0.04	0.02	1.00	0.03
HASAR_RED_KAZA_SAYI	-0.01	0.05	-0.00	0.08	0.14	-0.02	-0.02	0.06	0.41	0.47	0.03	0.54	0.01	0.17	0.03	1.00

	HUKUK_DOSYASI_ILISKISI	HASAR_OTO_KISISEL_SEBEP_SAYI	HASAR_OTO_DISKAYNAK_SEBEP_FLAG	HASAR_OTO_DISKAYNAK_SEBEP_SAYI	HASAR_OTODISI_GENEL_SEBEP_FLAG	HASAR_OTODISI_GENEL_SEBEP_SAYI	HASAR_OTDSI_KISISEL_SBP_FLAG	HASAR_OTDSI_KISISEL_SBP_SAYI	HASAR_OTDSI_DSKYNK_SBP_FLAG	HASAR_OTDSI_DSKYNK_SBP_SAYI	HASAR_OTDSI_DHLSU_SBP_FLAG	HASAR_OTDSI_DHLSU_SBP_SAYI	YUKLU_AGIRLIK	BEYGIR_GUCU	SILINDIR_HACMI
HUKUK_DOSYASI_ILISKISI	1.00	-0.01	-0.01	-0.00	-0.01	0.01	0.00	-0.00	-0.00	-0.00	-0.01	0.01	0.02	-0.00	-0.01
HASAR_OTO_KISISEL_SEBEP_SAYI	-0.01	1.00	0.41	0.52	0.33	0.04	0.54	0.55	0.05	0.01	0.14	0.01	-0.11	0.08	0.16
HASAR_OTO_DISKAYNAK_SEBEP_FLAG	-0.01	0.41	1.00	0.24	0.30	0.10	0.22	0.12	0.12	0.06	0.17	0.07	-0.02	0.10	0.12
HASAR_OTO_DISKAYNAK_SEBEP_SAYI	-0.00	0.52	0.24	1.00	0.23	0.01	0.44	0.10	0.00	0.00	0.00	-0.00	-0.02	0.02	-0.00
HASAR_OTODISI_GENEL_SEBEP_FLAG	-0.01	0.33	0.30	0.23	1.00	0.21	0.41	0.20	0.27	0.13	0.36	0.14	-0.01	0.10	0.12
HASAR_OTODISI_GENEL_SEBEP_SAYI	0.01	0.04	0.10	0.01	0.21	1.00	0.22	0.09	0.28	0.21	0.28	0.73	0.01	0.01	0.02
HASAR_OTDSI_KISISEL_SBP_FLAG	0.00	0.54	0.22	0.44	0.41	0.22	1.00	0.46	0.20	0.07	0.16	0.12	-0.08	0.09	0.16
HASAR_OTDSI_KISISEL_SBP_SAYI	-0.00	0.55	0.12	0.10	0.20	0.09	0.46	1.00	0.08	0.03	0.12	0.03	-0.07	0.06	0.12
HASAR_OTDSI_DSKYNK_SBP_FLAG	-0.00	0.05	0.12	0.00	0.27	0.28	0.20	0.08	1.00	0.42	0.35	0.28	0.01	0.01	0.02
HASAR_OTDSI_DSKYNK_SBP_SAYI	-0.00	0.01	0.06	0.00	0.13	0.21	0.07	0.03	0.42	1.00	0.20	0.54	0.02	0.01	0.01
HASAR_OTDSI_DHLSU_SBP_FLAG	-0.01	0.14	0.17	0.00	0.36	0.28	0.16	0.12	0.35	0.20	1.00	0.33	0.03	0.03	0.01
HASAR_OTDSI_DHLSU_SBP_SAYI	0.01	0.01	0.07	-0.00	0.14	0.73	0.12	0.03	0.28	0.54	0.33	1.00	0.03	0.01	-0.01
YUKLU_AGIRLIK	0.02	-0.11	-0.02	-0.02	-0.01	0.01	-0.08	-0.07	0.01	0.02	0.03	0.03	1.00	0.28	-0.49
BEYGIR_GUCU	-0.00	0.08	0.10	0.02	0.10	0.01	0.09	0.06	0.01	0.01	0.03	0.01	0.28	1.00	0.14
SILINDIR_HACMI	-0.01	0.16	0.12	-0.00	0.12	0.02	0.16	0.12	0.02	0.01	0.01	-0.01	-0.49	0.14	1.00

Figure 5. Correlation coefficient results

When the figures are examined, the most important factor variables with high correlation with each other should be removed from the data set because these two variables have the same effect on the target variable.

Conclusion

In this study, in summary, a classification problem was discussed. Before running machine learning algorithms, by looking at the significance levels of the variables in the data set, it was observed using more than one method which variables would be significant for the model and which variables had a common effect on the model. There are multiple variable selection algorithms in the literature [16].

Recommendations

In the study, studies were made on some of the variable selection algorithms found in the literature according to the data set, and the results were shared. Other algorithms in the literature can be tried in study-specific or further studies. Both categorical and numerical variables can be examined separately, and a variable selection can be made according to the data type.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This work would not have been possible without the financial support of Anadolu Insurance. I am grateful to Data Management and Analytics collegues, and Data Analysis and Management Reporting team, and all of those with whom we have had the pleasure to work during this and other related projects.

* This article was presented as oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Marmaris/Turkey on April 27-30, 2023.

References

Akar B. (2021), Müşteriye özel fiyat tahmin çalışması, *YBS Ansiklopedi, 9, 13 – 29.*

Aker Y. (2022), Comparison of PCA and RFE-RF algorithm in bankruptcy prediction, *Gümüşhane University Journal of Social Sciences Institue. 13,* 1001 – 1008.

Garcia Y., Garcia B., Gomez M., Fernandez B., & Garcia C. J. (2017), Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data, *BMC Medical Informatics and Decision Making, 17,* 34 – 38.

Gregorutti B., & Michel B. (2017), Correlation and variable importance in random forests, *Springer link, 27, 659 – 678.*

İlhan A, & Sarı M. H. (2016), Marmara gölü'ndeki (Manisa) vimba vimba (Eğrez) Populasyonunun bazı biyolojik özellikleri. *Jurnal of Limnology and Freshwater Fisheries Research, 2,* 59 – 65

Kaya H., & Köymen K. (2008), Veri madenciliği kavramı ve uygulama alanları. *Fırat Üniversitesi Doğu Araştırmaları Dergisi, 2,* 159 – 164

Khan A, & Cotton C. (2023), Efficient attack detection in IoT devices using feature engineering-less machine learning. *https://arxiv.org/abs/2301.03532, 14,* 47- 64.

Kirisci M. (2022), Correlation coefficients of fermatean fuzzy sets with a medical application, *Journal of Mathematical Sciences and Modelling. 1,* 16 – 23.

Kozak J, Kaine K, & Juszczuk P. (2020), Permutation entropy as a measure of ınformation gain/loss in the different symbolic descriptions of financial data. *https://www.mdpi.com/1099-4300/22/3/330#, 22,* 330 – 341

Li X., & Liu J. (2017), Automatic essay scoring based on coh-metrix feature selection for Chinese English learners, *Emerging Technologies for Education, 2*, 382 – 393.

Santos D. O. V., Verspoor M., & Nerbonne J. (2012), Identifying important factors in essay grading using machine learning. *Oxford Handbook of Applied Linguistics.* 1 – 15.

Sharma S., & Goyal A. (2020), Automated essay grading: An emprical analysis of ensemble learning techniques, *Computational Methods and Data Engineering.* 343 – 362.

Uğurlu M, Doğru İ, & Arslan S.R. (2023). Karanlık ağ trafiğinin makine öğrenmesi yöntemleri kullanılarak tespiti ve sınıflandırılması. *Jurnal of the Faculty of Engineering and Architecture of Gazi Universty, 38,* 1737-1746.

Zhang S., Zhu F., Qianhao Y, & Xiaoyue Z. (2021) Identifying DNA-binding proteins based on multi-features and LASSO feature selection. *Wiley Online Library Biopolymers 112,* 17 – 28

Author Information

Vedat Güneş
Anadolu Anonim Türk Sigorta Şirketi
Rüzgarlıbahçe, Çam Pınarı Sok. No:6, 34805
Beykoz/İstanbul, Türkiye
Contact e-mail: *vgunes@anadolusigorta.com.tr*

Serkan Kırca
Anadolu Anonim Türk Sigorta Şirketi
Rüzgarlıbahçe, Çam Pınarı Sok. No:6, 34805
Beykoz/İstanbul, Türkiye

Hasan Ersan Yağcı
Muğla Sıtkı Koçman Üniversitesi
Fen Bilimleri Enstitüsü, Yapay Zeka ABD, 48000
Menteşe/Muğla, Türkiye

Nida Gökçe Narin
Muğla Sıtkı Koçman Üniversitesi
İstatistik Bölümü, Yapay Zeka Laboratuvarı, 48000
Menteşe/Muğla, Türkiye

To cite this article:

Gunes, V., Kirca, S., Yagci, H.E., & Narin, N.G. (2023). Variable selection with machine learning in the legalization process for traffic insurance. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 22,* 237-246