Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Special Issue: Artificial Intelligence

Scientific Paper

Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210140/2022

G-SOJA - WEBSITE WITH PREDICTION ON SOYBEAN CLASSIFICATION USING MACHINE LEARNING

Daniela C. de Oliveira[1*], Uender C. Barbosa[2], Alcídia C. R. O. Bergland[2], Osvaldo Resende[1], Daniel E. C. de Oliveira[1]

[1*]Corresponding author. Instituto Federal Goiano/Rio Verde - Goiás, Brasil.
E-mail: danielacaboliveira@gmail.com | ORCID ID: https://orcid.org/0000-0002-9647-933X

KEYWORDS

standard, non-standard, decision-making.

ABSTRACT

This study is dedicated to the development of a methodology based on supervised machine learning for soybean classification and justified as technological innovation to predict whether soybean classification is in the standard or non-standard established by normative instruction No. 11/2007 of the Ministry of Agriculture, Livestock, and Food Supply (MAPA). This study aimed to develop a website using supervised machine learning to classify soybeans, providing an assertive decision-making process in real-time. A technological tool was created to assist the farmer and the storage unit in the classification of soybeans, considering the perceived reality and potential instruments consistent with the reality of the area. Therefore, a website in Python language was created using the Pandas, Pandas Profiling, Seaborn, Matplotlib, NumPy, Scikit-learn, PyCaret, and Streamlit libraries. In the end, the system could predict whether the soybean is in the standard or non-standard established by the soybean classification normative. In this sense, the results showed the robustness and precision of the proposed new methodology.

INTRODUCTION

Grain classification is regulated in accordance with the legislation and instructional norms established by the Ministry of Agriculture, Livestock, and Food Supply (MAPA), which designates classification rules for each grain. Therefore, specific analyses are carried out on samples to assess the qualitative aspects of the product and compare them with the official standards established by MAPA, which are finally categorized by Group, Class, and Type of grain.

According to Ferraz & Pinto (2017), the technologies are used "[…] for the simple consultation of weather conditions or agricultural commodity prices and even in the property accounting and use of precision machines." Thus, the available tools assist in the decision process, with the adequate allocation of resources and possibilities for reducing risks and increasing profits.

In this context, the technological massification aimed at agricultural activities, which include the different segments that make up the production chains, is a phenomenon characterized by a set of technologies, such as artificial intelligence, data science, big data, internet of things, and machine learning. The technology developed for the field aims to increase yield, efficiency, and profit.

The deployment of technology in the field requires the interconnection of elements, data collection and processing, storage, and mainly efficient use. Disconnected data requires the use of detection patterns, data mining, artificial intelligence, and deep learning, generating relevant information for decision-making and application of the best agricultural strategy.

Machine learning is the branch of artificial intelligence that has great prominence. Santos et al. (2019) stated that machine learning is one of the most studied areas in artificial intelligence and data classification. This technique allows the identification of patterns based on previous cases and experiments, as it happens with human intelligence.

[1] Instituto Federal Goiano/ Rio Verde - GO, Brasil.
[2] Instituto Federal Goiano/ Iporá - GO, Brasil.

Area Editor: Gizele Ingrid Gadotti
Received in: 8-19-2021
Accepted in: 11-8-2021

According to Rolim et al. (2017), machine learning is one of the areas of artificial intelligence, which is defined as systems capable of acquiring knowledge from data. It can be divided into (i) supervised learning, which comprises the relationship between the provided inputs and outputs to classify or label a given instance in a set of pre-defined categories; and (ii) unsupervised learning, which groups elements with similar characteristics.

Jha et al. (2019) argued that agriculture faces challenge every day and the essential problems faced by farmers range from sowing to harvesting crops, and artificial intelligence and machine learning can penetrate all these categories.

Supervised machine learning is well present in the agricultural sector. According to Jha et al. (2019), neural networks have been incorporated into the agricultural sector due to their advantages over traditional systems, with the benefit of prevention based on parallel reasoning. The authors also claimed that the expert system PRITHVI, based on fuzzy logic, was developed to help farmers increase soybean production.

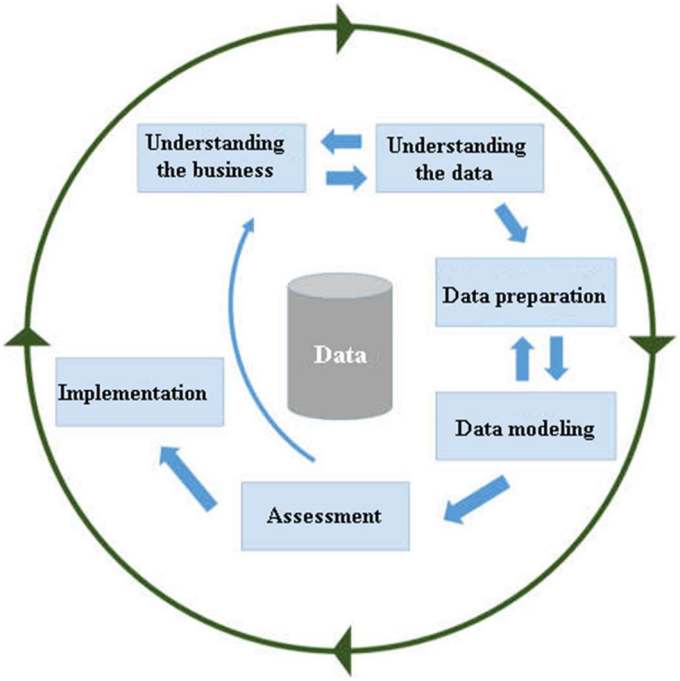
Moreover, Tu et al. (2021) also reported a low-cost, efficient, and non-destructive method to detect the corn seed variety based on image processing with deep learning.

Thus, this work proposal portrays an innovative supervised machine learning methodology aiming to develop a website to predict the standard and non-standard of soybean grain classification, providing assertive decision-making based on the MAPA normative instruction and real-time.

MATERIAL AND METHODS

The study research model is characterized as exploratory. The exploratory research seeks, through a bibliographic survey, interviews, and studies with experts, higher knowledge on the concepts related to the used technologies, communication in the rural environment in terms of opportunities or challenges, and legal parameters that regulate soybean classification in Brazil and their application. The qualitative technique was adopted as the research approach.

This study used the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to manage the project and development of the web application (Figure 1), being able to predict whether the classification of soybean grains is in the standard or non-standard, as established by Normative Instruction No. 11/2007 of MAPA.



Source: Prepared by the authors (2021).

FIGURE 1. Simplified design representation.

Understanding the business

The activities of receiving and storing soybean were monitored in a grain storage unit in the microregion of Iporá, Goiás, Brazil, to understand the business, with the case study being carried out to implement the technique. According to Barbosa et al. (2020), the classification process can be synthesized by the sequence of the following steps: sampling, homogenization, quartering, determination of foreign matter and impurity, determination of water content (moisture) of grains, determination of Group/Class/Type (e.g., soybean, corn, and bean), and report issuance, with the farmer's remuneration based on the sampling. This diagnosis was carried out in loco, following the routine of receiving and classifying soybean grains in the storage unit.

The problem involves the need to determine, through soybean classification data, whether the lot of grains fits the standard or non-standard, as designated by Normative Instruction No. 11/2007 of MAPA.

Understanding the data

The dataset was obtained from the mentioned case study and had an XLS format, being composed of 851 rows and 22 variables (columns), totaling 18,722 data. The dataset was converted into comma-separated values (CSV) and then an exploratory data analysis, also known as descriptive statistics, which corresponds to summarizing,

organizing, and interpreting the collected data, was performed. The columns of the initial dataset and their descriptions are shown below:

- Ticket No.: classification ticket number.
- Customer: customer name.
- Sample mass (g): weight of the sample mass in grams.
- Moisture (%): water content (% wb).
- Moisture (kg): moisture discount in kilograms.
- Impurity (%): percentage of impurity.
- Impurity discount (kg): impurity discount in kilograms.
- Green (%): percentage of greenish grains.
- Green discount (kg): discount for greenish grains in kilograms.
- Broken (%): percentage of broken/split/smashed grains.
- Broken discount (kg): discount for broken/split/smashed grains in kilograms.
- Damaged (%): percentage of damaged grains.
- Damaged discount (kg): discount for damaged grains in kilograms.
- Partial net (kg): net weight of cargo without discount in kilograms.
- Total discounts (kg): total discounts in kilograms.
- Net weight (kg): net weight of the cargo with discounts in kilograms.

The previous data analysis showed that the attributes were not standardized in grams. Charts allowed the visualization of the distributions of attributes that had outliers and missing data.

Data mining

Outliers and missing data, which compromise the performance of machine learning, were removed at this stage. The noises, which hinder pattern recognition, were removed for data mining.

In this context, four columns were added: impurity, greenish, broken, and damaged, and their values were converted into grams. Subsequently, the comma was changed to a period, leaving one decimal place after the period. An object-type column named target was also created, being the target variable of the machine learning model, with standard and non-standard values. Standard and non-standard values were created according to the column discounts (kg) as a reference, that is, the soybean was classified as non-standard in cases in which there were discounts and standard in cases in which there were no discounts.

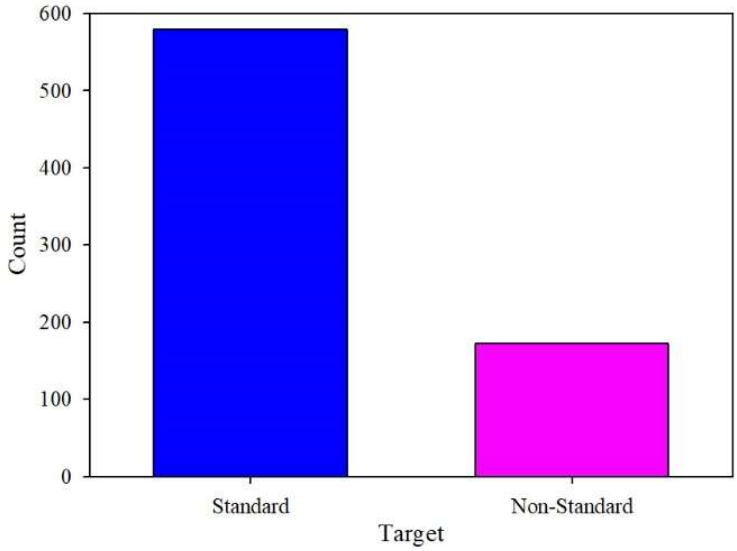
The columns Ticket No. and Customer were excluded from the dataset, as they are not necessary variables to predict the soybean classification. The columns Moisture (kg), Impurity (%), Impurity discount (kg), Green (%), Green discount (kg), Broken/split/smashed (%), Broken discount (kg), Damaged (%), Damaged discount (kg), Partial net (kg), Total discounts (kg), Net weight (kg) were also excluded, as they were replaced by the variables created from them. Figure 2 shows the new dataset, composed of 751 rows and 7 columns, totaling 5,257 data.

	A	B	C	D	E	F	G
1	sample_mass	moisture	impurities	green	broken	damaged	target
2	500.0	26.3	11.5	0.0	0.0	192.7	non_standard
3	500.0	25.2	11.5	0.0	0.0	176.7	non_standard
4	500.0	32.2	13.0	0.0	0.0	172.7	non_standard
5	500.0	34.6	20.0	0.0	0.0	172.7	non_standard
6	500.0	26.7	11.0	0.0	0.0	172.7	non_standard
7	500.0	26.6	15.0	0.0	0.0	170.1	non_standard
8	500.0	28.5	11.0	0.0	0.0	159.4	non_standard
9	500.0	21.3	22.0	0.0	0.0	132.9	non_standard
10	500.0	14.6	10.0	0.0	0.0	132.9	non_standard
11	500.0	33.1	16.0	0.0	0.0	130.2	non_standard
12	500.0	29.0	13.0	0.0	0.0	127.6	non_standard
13	500.0	21.7	16.0	0.0	0.0	127.6	non_standard
14	500.0	23.5	16.0	0.0	0.0	106.3	non_standard
15	500.0	23.4	17.0	0.0	0.0	106.3	non_standard
16	500.0	13.9	5.0	0.0	0.0	106.3	standard
17	500.0	13.9	5.0	0.0	0.0	106.3	standard
18	500.0	13.5	3.5	0.0	0.0	106.3	standard
19	500.0	13.5	3.5	0.0	0.0	106.3	standard
20	500.0	13.4	3.5	0.0	0.0	106.3	standard

Source: Prepared by the authors (2021).

FIGURE 2. Dataset after preprocessing.

Data mining for soybean classification allowed concluding that the dataset was composed of most of the data corresponding to the classification of standard soybeans (Figure 3).



Source: Prepared by the authors (2021).

FIGURE 3. Count of standard and non-standard values in the dataset.

The grain classification must comply with the following criteria to determine that the soybean classification is classified as standard: water content (moisture) = 14%; impurity and foreign matter = 1%; damaged = 8%; greenish = 8%; and broken, split, and smashed = 30%, as determined by the normative instruction of MAPA.

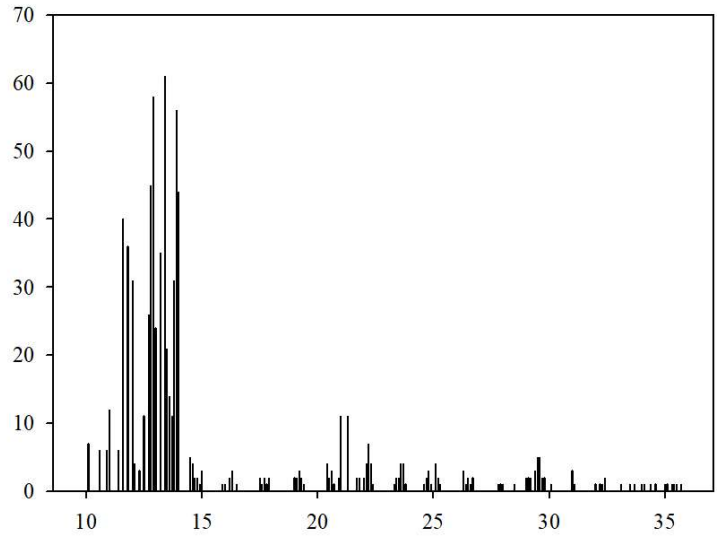
Subsequently, the statistical analysis of the dataset was performed by transforming the target column into binary, turning them into the columns target_standard and target_non_standard. Figure 4 shows the statistical analysis of the dataset, with the mean, standard deviation, minimum, quartiles (1st, 2nd, and 3rd), and maximum values for each column of the dataset.

	sample_mass	moisture	impurities	green	broken	damaged	target_non_standard	target_standard
count	751.0	751.000000	751.000000	751.000000	751.000000	751.000000	751.000000	751.000000
mean	500.0	15.280692	6.248336	0.721704	1.850866	57.630626	0.229028	0.770972
std	0.0	5.366410	4.868735	5.698226	14.865051	29.511149	0.420487	0.420487
min	500.0	10.100000	0.500000	0.000000	0.000000	13.300000	0.000000	0.000000
25%	500.0	12.750000	4.000000	0.000000	0.000000	39.900000	0.000000	1.000000
50%	500.0	13.400000	5.000000	0.000000	0.000000	53.100000	0.000000	1.000000
75%	500.0	14.000000	5.000000	0.000000	0.000000	79.700000	0.000000	1.000000
max	500.0	35.700000	23.000000	75.000000	165.000000	192.700000	1.000000	1.000000

Source: Prepared by the authors (2021).

FIGURE 4. Statistical analysis of the dataset.

Figure 5 shows another chart with detailed information on the dataset, for instance, the variable moisture with distinct, null, mean, maximum, and minimum values, among others. The histogram allowed verifying the data distribution.



Source: Prepared by the authors (2021).
FIGURE 5. Moisture column in detail.

Data modeling

The following technologies were used for data modeling: GitHub tool for project version control, Python programming language for the development of the machine learning model and API of soybean classification. The following libraries for Python programming language were also used: Pandas, Pandas Profiling, Seaborn, Matplotlib, NumPy, Scikit-learn, PyCaret, and Streamlit.

SQlite3, a module that provides an interface to SQlite[3] databases in Structured Query Language (SQL) adapted to the DB-API 2.0 specification, was chosen as the database. Anaconda, an essential package manager for data science and machine learning, was used. The Jupyter Notebook tool was used as an environment for the development, training, testing, and assessment of the results of the machine learning model. The Integrated Development Environment (IDE) Visual Studio Code was used.

Docker was used to deploying the API REST in production by creating a container with all dependencies. Docker is an open-source project, independent of languages and databases, running them inside containers. A container is a grouping of applications together with their dependencies, which share the operating system kernel. Kubernetes, an open-source container orchestration tool that automates the deployment, scheduling, and management of applications in containers, was chosen.

The following features were defined for dataset training: sample mass, moisture, impurities, greenish, broken/split/smashed, and damaged and the target variable. The training set was defined as the data presented to the machine learning algorithm to create the model with 70% of the data. The testing set is presented to the model after its creation, simulating actual predictions that the model has made, thus allowing the actual performance to be verified, that is, 30% of the data. The training and testing sets were separated, as shown in Figure 6.

```
: X_train, X_test, y_train, y_test = train_test_split(x,y,test_size =0.3)
```

Source: Prepared by the authors (2021).
FIGURE 6. Separation of training and testing sets.

Then, the environment in PyCaret was initialized through the setup function and the transformation pipeline was created, preparing the data for modeling and deployment, as shown in Figure 7.

[3] Library that uses an open-source, relational, cross-platform, client-server framework-independent SQL database engine.

```
s = setup(train
          ,target = 'target'
          ,numeric_features = [ 'sample_mass'
                               ,'moisture'
                               ,'impurities'
                               ,'green'
                               ,'broken'
                               ,'damaged']
          ,log_experiment = True
          ,experiment_name = 'exp'
         )
```

Source: Prepared by the authors (2021).

FIGURE 7. PyCaret initialization.

The PyCaret algorithm run the pre-processing tasks when the setup function was performed to train the model with all the algorithms present in the library in the classification module.

The information grid was printed once the data were checked and the configuration performed, as shown in Figure 8.

	Description	Value
0	session_id	1776
1	Target	target
2	Target Type	Binary
3	Label Encoded	non_standard: 0, standard: 1
4	Original Data	(751, 7)
5	Missing Values	False
6	Numeric Features	6
7	Categorical Features	0
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(525, 6)
12	Transformed Test Set	(226, 6)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	True

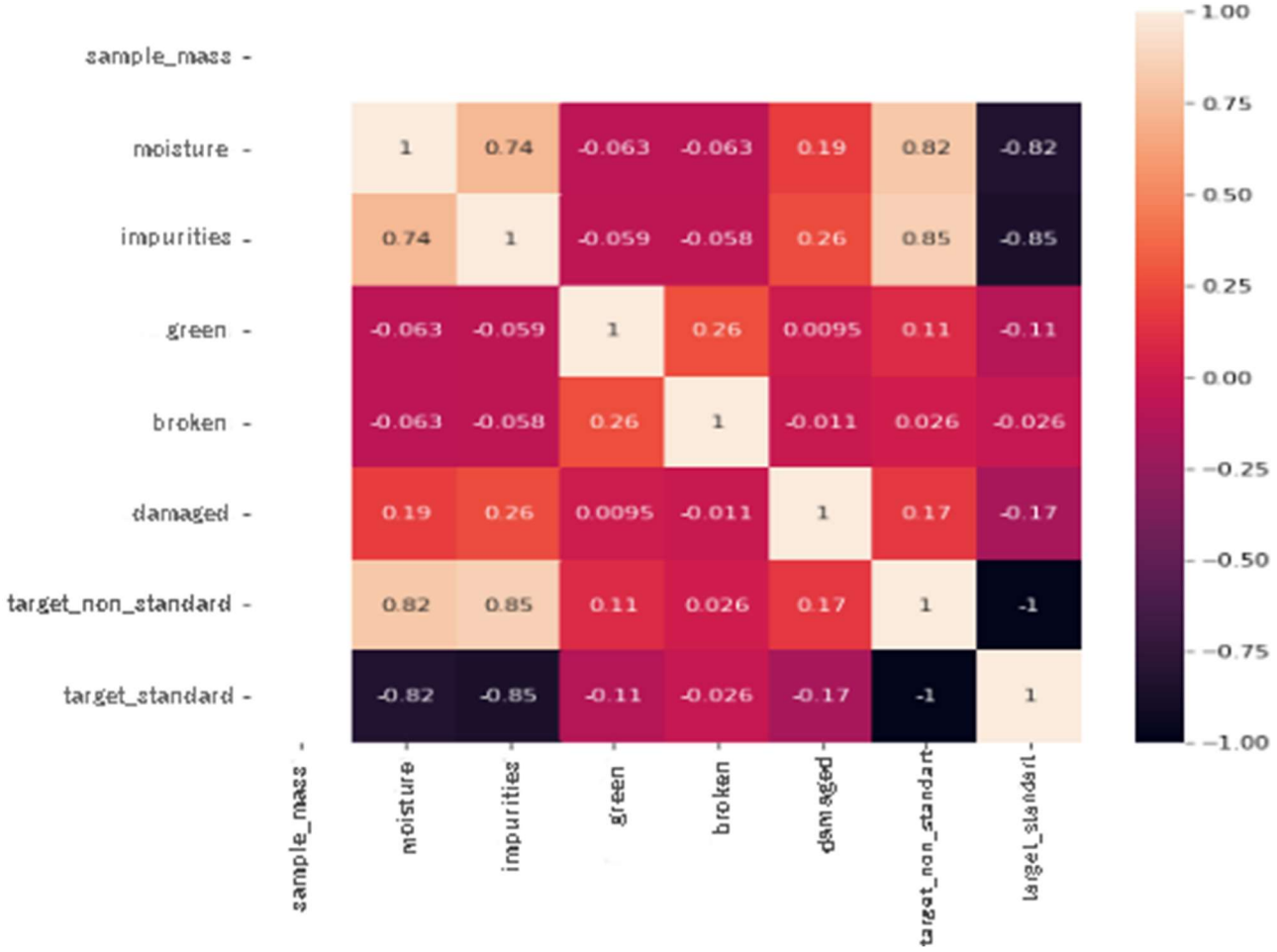
Source: Prepared by the authors (2021).

FIGURE 8. Grid with important information.

The grid portrayed the following information: Target Type – binary, the target variable was automatically detected and displayed; Label Encoded – the target variable was a 'standard' or 'non_standard' object type, and the label was automatically encoded in 0 and 1, displaying the mapping (non_standard:0, standard:1) for reference; Original Data – original form of the dataset in this experiment (751, 7), that is, 751 samples and seven resources, including the target column; Missing Values – there are no missing values in the dataset; Numeric Features – six out of the seven features are inferred as numeric; Categorical Features – there are no categorical features; Transformed Train Set – displays the transformed training set, in which the original form was (751, 7) and has been transformed into (525, 6) for the training set; and Transformed Test Set – displays the transformed testing set, that is, 226 samples in the testing set. This division of the training and testing datasets was based on the default value of 70/30.

RESULTS AND DISCUSSIONS

The statistical analysis of the data was obtained through Pearson's matrix as a result of data mining. In turn, Pearson's correlation matrix (r) signals the correlation between variables through the color intensity, that is, the variables have a higher correlation when the color intensity is light or close to −1 and, otherwise, a lower correlation (Figure 9). According to Miot (2008), Pearson's correlation coefficient is a statistical test that explores the intensity and direction of mutual behavior between variables. This coefficient can only assume values between −1 and 1.



Source: Prepared by the authors (2021).

FIGURE 9. Correlation between variables of the dataset.

In this sense, the variables impurities and target_non_standard are correlated, as well as the variables moisture and target_non_standard. Thus, these variables are associated with each other. On the other hand, the variables impurity and target_standard are not correlated, as well as the variables moisture and target_standard, that is, the variables have no association with each other.

Data modeling and the PyCaret library through the compare_models function allowed obtaining the score of the 15 best-supervised machine learning algorithms with stratified cross-validation to assess the algorithms, that is, 15 algorithms were compared with 5-time cross-validation (Figure 10). The data were not balanced, as the purpose of the proposal was to use real data from the storage unit of the microregion of Iporá, Goiás, and data mining was the only treatment performed.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.9924	0.9998	0.9855	0.9924	0.9923	0.9774	0.9775	0.1460
dt	Decision Tree Classifier	0.9905	0.9779	0.9779	0.9907	0.9903	0.9710	0.9717	0.0120
rf	Random Forest Classifier	0.9905	0.9951	0.9779	0.9907	0.9903	0.9710	0.9717	0.1700
ada	Ada Boost Classifier	0.9905	0.9783	0.9779	0.9907	0.9903	0.9710	0.9717	0.0720
gbc	Gradient Boosting Classifier	0.9905	0.9774	0.9779	0.9907	0.9903	0.9710	0.9717	0.0520
catboost	CatBoost Classifier	0.9905	0.9993	0.9779	0.9907	0.9903	0.9710	0.9717	0.7420
lightgbm	Light Gradient Boosting Machine	0.9867	0.9956	0.9721	0.9871	0.9864	0.9592	0.9604	0.4100
xgboost	Extreme Gradient Boosting	0.9848	0.9971	0.9646	0.9852	0.9844	0.9530	0.9546	0.1180
lr	Logistic Regression	0.9829	0.9972	0.9699	0.9830	0.9828	0.9490	0.9495	1.2540
nb	Naive Bayes	0.9810	0.9958	0.9749	0.9816	0.9810	0.9442	0.9450	0.0120
ridge	Ridge Classifier	0.9810	0.0000	0.9622	0.9815	0.9807	0.9424	0.9439	0.0160
lda	Linear Discriminant Analysis	0.9810	0.9972	0.9622	0.9815	0.9807	0.9424	0.9439	0.0160
knn	K Neighbors Classifier	0.9790	0.9856	0.9518	0.9797	0.9786	0.9357	0.9380	1.2240
svm	SVM - Linear Kernel	0.9733	0.0000	0.9446	0.9741	0.9727	0.9178	0.9207	0.0180
qda	Quadratic Discriminant Analysis	0.2171	0.0000	0.5000	0.0472	0.0775	0.0000	0.0000	0.0280

Source: Prepared by the authors (2021).
FIGURE 10. Comparison of performance measures of the applied classification algorithms.

The used metrics consisted of Accuracy, which indicated the model performance; Area Under the Curve (AUC), which provided the performance measure of the classification limits; Recall, which measured the amount of disapproved comments that the system approved; Precision indicated the positive class ratings, that is, how many are correct; F1 Score indicated the harmonic mean calculated based on precision and recall; Kappa, which measured inter-rater reliability; and Matthews Correlation Coefficient (MCC), which measured the quality of binary classifiers. The Extra Trees Classifier for supervised machine learning obtained higher values for the metrics: accuracy, AUC, recall, precision, F1 Score, Kappa, and MCC.

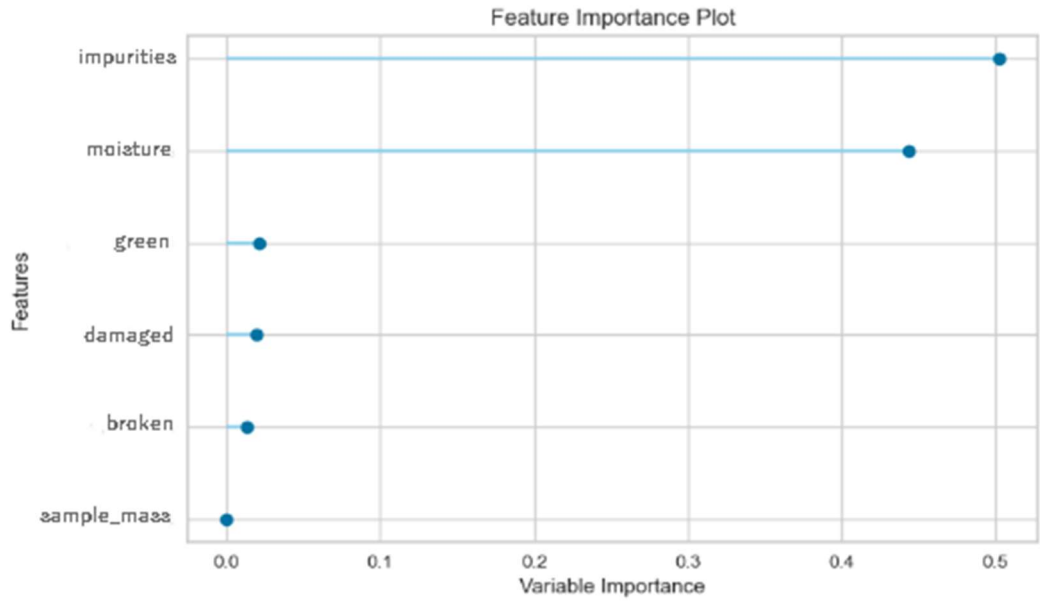
The Extra Trees Classifier algorithm obtained the highest rating for supervised machine learning metrics, which are: 99.24% accuracy, 99.98% AUC, 98.55% recall, 99.24% precision, 49.99% F1 Score, 97.74% Kappa, and 45.72% MCC. The means of the model evaluation metrics were determined by the mean and standard deviation (Figure 11).

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9905	1.0000	0.9773	0.9906	0.9904	0.9708	0.9712
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.9810	0.9992	0.9722	0.9810	0.9810	0.9443	0.9443
3	0.9905	1.0000	0.9783	0.9906	0.9904	0.9717	0.9721
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Mean	0.9924	0.9998	0.9855	0.9924	0.9923	0.9774	0.9775
SD	0.0071	0.0003	0.0120	0.0071	0.0071	0.0209	0.0209

Source: Prepared by the authors (2021).
FIGURE 11. Mean of model evaluation metrics.

The metrics presented above allowed concluding that the supervised machine learning algorithm Extra Trees Classifier had the best classification for soybean grains, followed by the Decision Tree and Random Forest Classifier algorithms.

Impurities and moisture were important variables for the supervised machine learning model, followed by greenish, damaged, broken, and sample_mass (Figure 12).



Source: Prepared by the authors (2021).

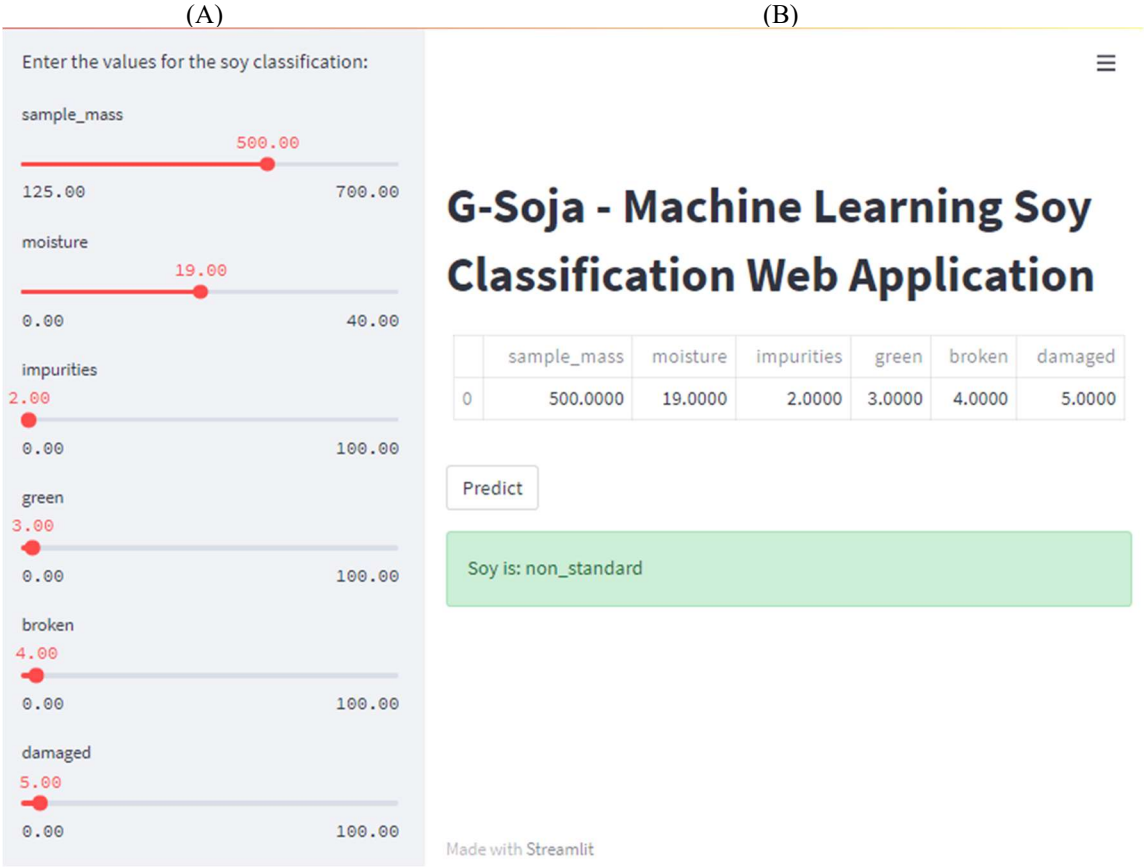
FIGURE 12. Determination of the importance of variables.

Therefore, the variable impurities directly influenced the supervised machine learning for soybean grain classification, followed by the variable moisture.

The website was developed after the completion of the supervised machine learning implementation, using the Rest API with the Streamlit library to provide the real-time running supervised machine learning model (Figure 13).

The website was made available in real-time to the producer and the storage unit, to which the purpose of entering data on the sample mass, moisture, impurities, greenish, broken, and damaged in the system was delegated.

(A) (B)



Source: Prepared by the authors (2021).

FIGURE 13. (A) Standard soybean classification screen; (B) website made available in real-time.

The website processes the prediction of soybean classification, allowing verifying whether the classification is in the standard or non-standard, as established by Normative Instruction No. 11/2007 of the Ministry of Agriculture, Livestock, and Food Supply (MAPA) and based on specialized literature.

The website has a computer program registration certificate with the Brazilian National Institute of Industrial Property entitled: G-Soja – Site Web para Classificação de Soja com Aprendizado de Máquina, issued on 02/02/2021, with code BR 512021000113-3.

Importantly, the data mining and modeling steps directly influenced the supervised machine learning, which presented a very relevant accuracy rate of 99.24%, making the result robust and precise.

The proposed methodology is considered innovative since there are no studies on the soybean classification in the literature. However, there are several applications of artificial intelligence for studies in the postharvest area.

Moslem et al. (2019) modeled the production of safflower seeds using machine learning algorithms, artificial neural networks, and multiple linear regression. Saffariha et al. (2020) performed the prediction of germination of *Salvia limbata* seeds under stress by applying artificial intelligence modeling techniques, such as the MLR and MLP algorithms. In turn, Medeiros et al. (2020) studied the classification of *Jatropha curcas* L. seeds using X-ray image analysis and machine learning. Finally, Spancerski & Santos (2021) proposed a model of LSTM recurrent neural networks for the prediction of rice yield in the state of Rio Grande do Sul, Brazil. The model presented adequate results for a short-term forecast.

CONCLUSIONS

The supervised machine learning proposed in this study with the development of a website for soybean grain classification based on intelligent computing techniques is an innovative methodology with efficient, reliable, robust, and precise results to perform the prediction of soybean classification according to normative instruction of MAPA.

ACKNOWLEDGEMENTS

The authors extend thanks to IF Goiano, CAPES, FAPEG, FINEP and CNPq for their financial support, which was indispensable to the execution of this study.

REFERENCES

Ali M Pycaret: About PyCaret. [S. l.]. (2020). Available: https://pycaret.org/about/. Accessed Jul 4, 2021.

Ali M Pycaret: Classification Module. Available: https://pycaret.org/classification1/. Accessed Jul 19, 2021.

Barbosa UC, Bergland ACRO, Oliveira DC, Oliveira DEC, Furquim MGD, Júnior JCS (2020). iGrãos: desenvolvimento de chatbot em redes sociais para classificação de soja para sojicultores. Pesquisa, Sociedade e Desenvolvimento, 9(10). DOI: 10.33448 / rsd-v9i10.8558.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. (2007). Mapa. Instrução Normativa Mapa nº 11, de 16 de maio de 2007. Estabelece o Regulamento Técnico da Soja, definindo o seu padrão oficial de classificação, com os requisitos de identidade e qualidade intrínseca e extrínseca, a amostragem e a marcação ou rotulagem. Brasília.

Ferraz, CO, Pinto WF (2017) Tecnologia da Informação para a Agropecuária: utilização ferramentas da tecnologia da informação no apoio a tomada de decisões em pequenas propriedades. Revista Eletrônica Competências Digitais para Agricultura Familiar 3(1):38-49, Tupã.

Jhan K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. Artificial Intelligence in Agriculture 1:1-12. DOI: https://doi.org/10.1016/j.aiia.2019.05.004

Medeiros AD de, Pinheiro DT, Xavier WA, Silva LJ da, Dias DCF dos S (2020) Quality classification of Jatropha curcas seeds using radiographic images and machine learning. Industrial Crops and Products 146:112–162. Doi: https://doi.org/10.1016/j.indcrop .2020.112162.

Miot HA (2018) Correlation analysis in clinical and experimental studies. Journal Vascular Brasileiro 17(4):275-279. DOI: https://doi.org/10.1590/1677-5449.174118

Moslem A, Younessi-Hmazekhanlu M, Ramazani SHR, Omidi AH (2019) Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (Carthamus tinctorius L.). Industrial Crops and Products 127:185-194. DOI: https://doi.org/10.1016/j.indcrop.2018.10.050.

Rolim VB, Mello RFL, Costa EB (2017) Utilização de técnicas de aprendizado de máquina para acompanhamento de fóruns educacionais. Revista Brasileira de Informática na Educação 25(3).

Saffariha M, Jahani A, Potter D (2020) Germination prediction of Salvia limbata seeds under ecological stress in protected areas: an artificial intelligence modeling approach. BMC Ecology 20:48. Doi: https://doi.org/10.1186/s12898-020-00316-4.

Santos KS, Júnior JRF, Wada DT, Tenório APM, Barbosa MHN, Marques PMA (2019) Inteligência Artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo a medicina de precisão. Radiologia Brasileira 52(6). DOI: https://doi.org/10.1590/0100-3984.2019.0049.

Scikit-Learn (2020) Getting started. a guide for people new to vision loss. VisionAware. Available: https://scikit-learn.org/stable/getting_started.html. Accessed Jun 10, 2021.

Spancerski JS, Santos JAA (2021) Previsão da produtividade de arroz: uma aplicação de redes neurais recorrentes LSTM. Revista Cereus 13:163-175. DOI: https://doi.org/10.18605/2175-7275.

Tu K, Wen S, Cheng Y, Zhang T (2021) A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968′using machine vision combined with deep learning. Computers and Electronic in Agriculture. 182. DOI: 10.1016 / j.compag.2021.106002