CERAL RATING PREDICTION USING MACHINE LEARNING

OCTOBER-16 ,2020

P. SABIHA *S. SWETHA* *B. SURYA NARAYA*

17MIS0064 *17MIS0084* *17MIS0035*

DESIGN PATTERNS

RESEARCH PAPER

Abstract

In our project we we use the ML algorithms like Linear regression, SVR, Decision Tree, Random forest methods for the algorithm to predict cereals rating. Prediction is the method of making statements about certain event whose actual results have not been observed. It seems to be an easy process but is not. It requires a lot of analysis on current and past outcomes in order to give timely and accurate timely predicted results. Providing a decent customer satisfaction has always been a enormous trouble for companies, especially for cereal manufacturers. The gathering of the customer reviews has become a huge importance for the market sales. Therefore, predicting potential customer ratings have huge impact to increase sales and decrease negative feedback. Rating of products that do not come onto the market, cannot be labeled via human supervision. Here, we investigated solving this problem via machine learning methods such as SVR, Linear Regression, Decision Trees, Random Forests.

KEYWORDS:

Linear regression, SVR (support vector regression), Decision Tree, Random forest. Artificial Neural Networks, Back propagation (BP), Forward Propagation, Radial basis function network (RBF), Fuzzy set, Raw data , Multi-layer perceptron (MLP)

INTRODUCTION

Before starting to understand the dataset that we choose to examine and research, very well, we should understand carefully what is data , why we need data science, how to use a data set and also how to convert sets of data to an understandable data circuits. Before learn about data science, first word that should known as data. Data is a set of gathered and translated information form to use for analysis. Second word is big data. Big data, both structured and unstructured, are large collections and traditional software techniques are very large to control and analysis. Third but not least word is machine learning. Machine learning is a part of artificial intelligence that provides and focus on the development and future of computer programs. The mean reason for machine learning, allow the computers learn by themselves without human intervention and adjust actions accordingly.

Generally, the thing that includes data collection that cannot be solved by hand is data science. Data science is a science that helps us to clearly understand and provide meaningful usage of a big data collection. It uses computer science, statistics and machine learning, visualization and human computer interactions to collect, clean, integrate, analyse, visualize and interact with data to create understandable

data products. In Data Science process, there are steps that should consider carefully. These steps can count as Obtaining data, pre-processing of data and editing of data, creating models, testing models, measuring the performance of model.

After getting information about where we can bring this project to life, we search data sets that should have proper for data analysing process. In order to extensive data set research, we found a dataset which has ingredients in cereal food product with brand names in it. These ingredients will be our data that we will examined. During examination of dataset give us opportunity to determine the problem in dataset as prediction of customer cereal rating. Most of the people in the earth who has opportunity to eat proper foods, mostly eating cereals at their breakfast. This thought give us our problem that can solve with this dataset. In recent years, seventy-seven different cereal brands has come up. These cereals has different ingredients percent and has different taste. In this dataset, according to customers ratings, cereals rated. We plan to see can we predict possible rating for cereals which has ingredient percentage knowledge.

Manufacturer of cereal can include different values as American Home Food Products, General Mills, Kellogg, Nabisco, Post, Quaker Oats and Ralston Purina. These values are shown with their first letter in relevant cell. When examining to type as feature, it can include two value: C for cold and H for hot. For Rest of the features are numerical. We found this dataset from kaggle website.

II. DATASET

A. Dataset Description Our dataset contains 77 different type of cereals. Each type of cereals have 16 features such as:

- name: cereal's name
- mfr: manufacturer of cereal
- type: type of cereal
- calories: calories per serving
- protein: amount of protein as gram in one serving
- fat: amount of fat as gram in one serving
- sodium: amount of protein as milligram in one serving
- fiber: amount of dietary fiber as gram in one serving
- carbo: amount of complex carbohydrates as gram in one serving
- sugars: amount of sugar as gram in one serving
- potass: amount of potassium as milligram in one serving
- vitamins: vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA recommended
- shelf: display shelf (1, 2, or 3, counting from the floor)
- weight: weight in ounces of one serving
- cups: number of cups in one serving

III. PREPROCESSING

A. Missing Data

The preliminary data set we had about 77 manufacturer with 16 features. However the data was slightly "porous", as described by one associate. Some of cereal examples have not all the 15 features described, especially at amount of carbohydrates, potassium and sugar columns. This data values marked by "-1".An amount of nutrition can not be negative, so we can assume these are missing values. We needed a full data set to begin training our model.

A method was devised to fill up negative values. First, we tried replace this cells with mean of the column which is relevant, and tested it. Then, we tried replace this cells with zero, and tested it also. When we analyzed the results we saw that scores was better when we replace it with zero. So, any features which have negative values were filled in the working data set with zero. After this filling process, every cereal name with an missing value was handled. This gave us a working data set with no dimension lost. No need to delete missing rows

B. Encoding

Although most features were described with a manageable and numeric value, some were described with a categorical value that needed to be converted or deleted. For example, the value entered for mfr, which was a text entry from a consistent list of name of manufacturers,for which there were 7 categories. To make the training easier we gave an integer 0 - 6 for Manufacturer of cereal. Another feature which binary encoded is the type of cereal(hot or cold). Given the numerical values, 1 for"hot", 0 for "cold"

C. Feature Selection

There are some correlations between features. We had to analyse this relation for understanding the dataand acquiring more consistent predictions. We used Random Forest algorithm for scaling the relation of other features between "rating". According the Fig.1, the most important, most correlated feature with the target feature (rating) is "sugar" , then "calories" come. It show us they are the key values. The least important feature is "type". Probably, reason of that is being same (95%) most of values are same in that feature. Fig.1 shows us the importance power with only one other feature, this is not enough alone.

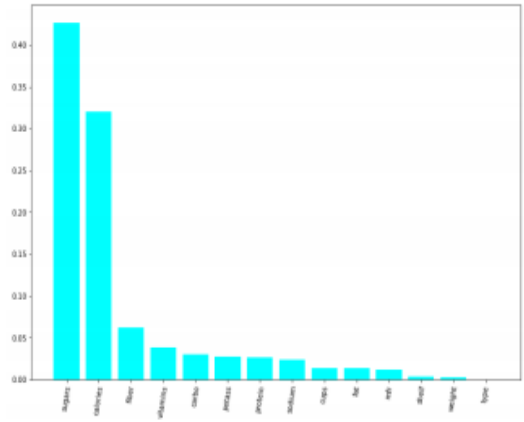


Fig.1 features important based on random forest

Pearson correlation heatmap; We used Pearson correlation heatmap (Figure 2) see the correlation of independent variables with each other. Also, heatmap told us, is the features positive or negative proportional correlated. According the Figure 2 cell value at intersection of "type" and "rating" is nearly zero. It show us their correlation equal to 0. The strongest correlated intersection between ranking is "sugar" which is equal to -0.76. According this value, amount of sugar is the important feature for determine the rating, and less sugar means better rating.

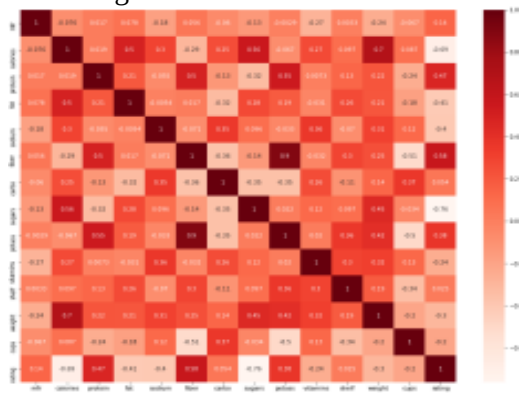


Fig.2 correlation heatmap

D. Deletion

Less is more. First of all, we deleted the name of cereal feature, because It is all the values are unique. So they are meaningless for our models. Secondly, we decided to remove the type of cereal feature. Because, according to our feature importance analysis type data contributes almost zero.

IV. METHODS &RESULTS:

A.Environment :

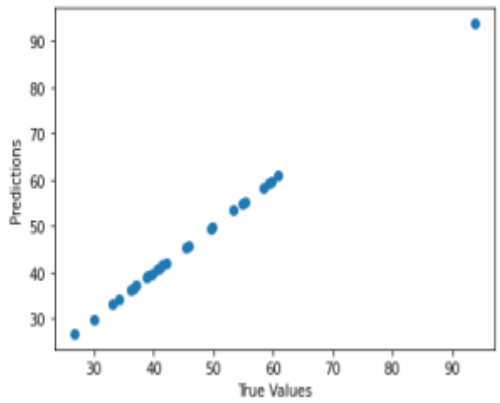
Here we used Python for implementation. We chose Python programming language for training and testing of our models with dataset. Python is a programming language, which used in data science because of algorithms and functions that used for data analysing and processing. As editor of Python, we use Anaconda and Spyder. They provide us lots of convenience. We can monetarize variable of data, output console and data frame. Spyder's variable explorer tool colourizing and sorting data frame values comprehensibly for us.

We separated less number(like 30 % or 20%) of the samples for testing, and separated rest(like 70% or 80%) of the for training. By using the same data for training and testing we can minimize the effects of data discrepancies and better understand the characteristics of the model. After a model has been processed by using the training set, we test model by making predictions against the test set. Because the data in the testing set already contains known values for the attribute that we want to predict, it is easy to determine that you want to predict, it is

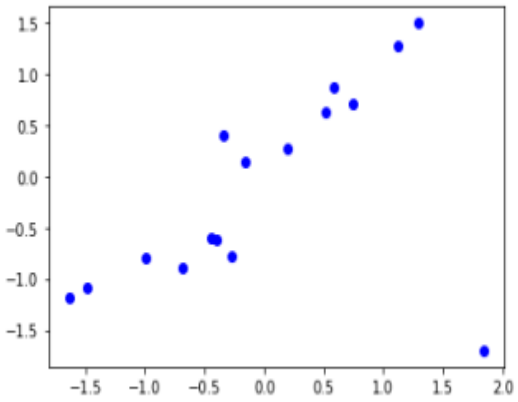
easy to determine whether the model's guesses are correct or not.

B.Machine learning methods :

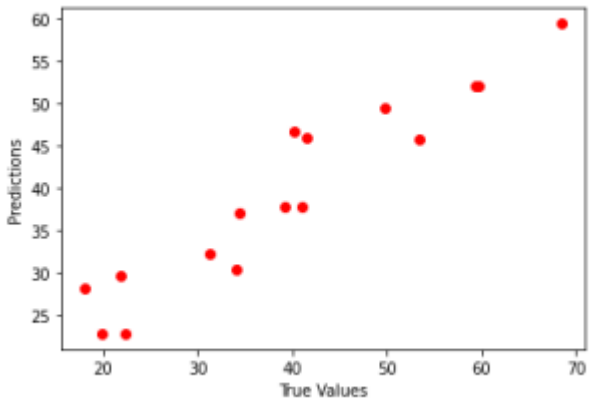
Linear Regression :
We began to analyze our data by implementing linear regression first. Linear regression is a relatively simple algorithm where the algorithm finds a line in higher dimensions such that the sum of the squared distance between the line and the data points is minimized. After having fit this line, the algorithm predicts the outcome of an unseen data point by plugging in the point's features to the line equation.
We separated 30% of the samples for testing, and separated rest of the 70 % for training.By using the same data for training and testing we can minimize the effects of data discrepancies and better understand the characteristics of the model.
The accuracy of the Linear Regression algorithm is '95.7%'.



SVR : (Support Vector Regression)
In simple regression we try to minimise the error rate. While in SVR we tried to fit the error within a certain threshold. We used it , because for assuming the algorithm can reduce the error. Unfortunately Support Vector Regression algorithms effecting easily from outliers while creating models. So, we need to data standardization. We used "StandardScaler" is that it will transform our data such that its distribution will have a mean value 0 and standard deviation
of 1.
We separated 20% of the samples for testing, and separate rest of them for training. We modeled 4 kernel methods which are the function used to map a lower dimensional data into a higher dimensional data for testing the results and use the best one.
The accuracy of the SVR algorithm is '97.4%'.

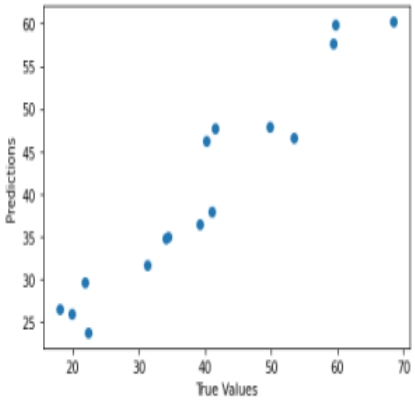


Decision Trees :
In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions.
In this model we use 20% for testing and 80 % for training.
The accuracy of the Decision Tree algorithm is '85.4%'.



Random Forests :
Random Forests also employ Bagging, Bootstrap Aggregation. When a new tree is trained, it is trained using a uniformly random sample of features from the original dataset with replacement. Sampling with replacement allows for a sample of training features to be

repeated. This is ”Bootstrapping”. Then, once all the trees are trained, they can be fed the testing data.Every tree willgive a different result for the test data, but the average for regression. Random Forests have many attractive qualities. Because Random Forests are essentially a collection of Decision Trees, they don’t need to be trained with normalized data, and can handle both continuous and categorical data simultaneously. Much like decision trees, they have white box qualities, as you can look at the trees that produce the results. One can also calculate the importance of variables by averaging the error difference as the values of a variable are permuted across all the trees. Bagging also makes Random Forests resilient against variance and overfitting. It’s likely that those are the reasons why it performed so well with our data.
The accuracy of the random forest algorithm is ’89.0’.



V. CONCLUSION :

Algorithms	Accuracy
Linear Regression	95.7%
Support Vector Regression	97.4%
Decision Trees	85.4%
Random Forests	89.0%

We used accuracy, precision and recall on the training and validation sets to evaluate the performance of each algorithm.
Although all algorithms have considerable predictions, The Decision Trees algorithm gave the least accuracy of 85.4%, with predictions on data points predict poorly to actual customer ratings. Although LR and SVR gives a very good accuracy of 95.7% and 97.4% against the validation data, with similar accuracy against the training data indicating no over fitting, SVR with Linear Kernel method is the "chef’s recommendation".
For the SVR, each kernel did not yield reasonable accuracy on the test data. Especially, sigmoid method have poor accuracy on the validation data. Using random forests did not improve the precision or recall. We could potentially increase the precision by collating a larger validation set. Our features have a smooth, nearly linear dependence on the covariates, then linear regression will model the dependence better than random forests,
which will basically approximate a linear curve with an ugly irregular step function. If the dependence is multivariate linear and smooth, with significant covariates producing the dependence, the fit performance of random forests can be expected to get worse and worse for larger and larger. RF has a much greater ability than a single decision tree to model linearity, since we are adding tree predictions together - but still, it’s just not very efficient to approximate a high-dimensional linear relationship with a series of step functions. We think this is the most likely theoretical explanation for RF underperforming linear regression.

VI. REFERENCES :

[1] Muthuswamy, Sujithra. (2018). Simulation and Prediction of Cereals Rating Using Neural Networks Model and Multiple Regressions.

https://www.researchgate.net/publication/329519200_Simulation_and_Prediction_of_Cereals_Rating_Using_Neural_Networks_Model_and_Multiple_Regressions

[2] Safa, M. & Samarasinghe, Sandhya & Nejat, M.. (2015). Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New

Zealand. Journal of Agricultural Science and Technology. 17. 791-803.

https://www.researchgate.net/publication/281961996_Prediction_of_Wheat_Production_Using_Artificial_Neural_Networks_and_Investigating_Indirect_Factors_Affecting_It_Case_Study_in_Canterbury_Province_New_Zealand

[3] Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389-10397.
http://www.scielo.org.pe/pdf/jefas/v21n41/a07v21n41.pdf

[4] T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 1990, pp. 1-6 vol.1, doi: 10.1109/IJCNN.1990.137535.

https://www.semanticscholar.org/paper/Stock-market-prediction-system-with-modular-neural-Kimoto-Asakawa/25b0d0316ece493899d74cfb98ce7b77dca8352e

[5] Gutierrez, S. S. M., Palacios, A. T., Pérez, S. L., & López, A. S. (2018). Strategies and technologies of cereal sector in the face of climatic variability. A literatura review. *Strategy, Technology & Society*, *6*(1).
http://www.ijsts.org/index.php/STS3/article/view/21

[6] Dyrmann, M., Skovsen, S., Laursen, M. S., & Jørgensen, R. N. (2018, June). Using a fully convolutional neural network for detecting locations of weeds in images from cereal fields. In *International Conference on Precision Agriculture*. International Society of Precision Agriculture.
https://pdfs.semanticscholar.org/9476/2a8f63bbda7cd5a5260b0afb6ed0e0e40d05.pdf

[7] D. Lien Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min and H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network," in IEEE Access, vol. 6, pp. 55392-55404, 2018, doi: 10.1109/ACCESS.2018.2868970.
https://www.researchgate.net/publication/327484245_Deep_Learning_Approach_for_Short-Term_Stock_Trends_Prediction_based_on_Two-stream_Gated_Recurrent_Unit_Network

[8] Chandra, A.L., Desai, S.V., Balasubramanian, V.N. et al. Active learning with point supervision for cost-effective panicle detection in cereal crops. Plant Methods 16, 34 (2020).

https://plantmethods.biomedcentral.com/articles/10.1186/s13007-020-00575-8

[9] Yuan, W., Wijewardane, N.K., Jenkins, S. *et al.* Early Prediction of Soybean Traits through Color and Texture Features of Canopy RGB Imagery. *Sci Rep* **9,** 14089 (2019).
https://www.nature.com/articles/s41598-019-50480-x

[10] Karimi, H., Skovsen, S., Dyrmann, M., & Nyholm Jørgensen, R. (2018). A Novel Locating System for Cereal Plant Stem Emerging Points' Detection Using a Convolutional Neural Network. *Sensors*, *18*(5), 1611.
https://www.mdpi.com/1424-8220/18/5/1611

[11] Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, *253*(3), 697-710.
https://ideas.repec.org/a/eee/ejores/v253y2016i3p697-710.html

[12] Saman Mohammadi, Maaroof Siosemarde, "Application of Artificial Neural Networks in Order to Predict Mahabad River Discharge" SCIENTIFIC RESEARCH, an Academic Publisher Adv., 2016.

https://www.researchgate.net/publication/303954123_Application_of_Artificial_Neural_Networks_in_Order_to_Predict_Mahabad_River_Discharge

[13] N. Gandhi, O. Petkar and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, 2016, pp. 105-110, doi: 10.1109/TIAR.2016.7801222.
https://ieeexplore.ieee.org/document/7801222

[14] Vivian P.S. Felipe, Martinho A.Silva, Bruno D.Valente..,"Using multiple regression, Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes", OXFORD ACADEMIC, Adv., 2015,vol.94.
https://www.researchgate.net/publication/273773636_Using_multiple_regression_Bayesian_networks_and_artificial_neural_networks_for_prediction_of_total_egg_production_in_European_quails_based_on_earlier_expressed_phenotypes

[15] LAL, S. B., & VARMA, S. P. (2014). "Identification of abiotic stress related cereal proteins based on their structural composition using back propagation networks". *J. Agroecol. Natural Resour. Manage*, *1*(4), 270-274.
https://www.krishisanskriti.org/vol_image/11Jun201510062507%20%20%20%20S%20B%20Lal%20%20%20%20270-274.pdf

[16] Golub, Alla, and James J. Binkley. "Consumer choice of breakfast cereals." JEL: consumer demand, healthy food, Breakfast Cereals, Departement of Agricultural Economics, Purdue University(2005).
https://www.researchgate.net/profile/Alla_Golub/publication/252821662_Consumer_Choice_of_Breakfast_Cereals/links/02e7e537bd0a459992000000.pdf

[17] Shiva Kumar,1 P. Srinivasa Pai,2 and B. R. Shrinivasa Rao, Radial-Basis-Function-Network-Based Prediction of Performance and Emission Characteristics in a Bio Diesel Engine Run on WCO Ester(2012).
https://www.hindawi.com/journals/aai/2012/610487/

[18] Salam, Abdul & Ariyaluran Habeeb, Riyaz Ahamed. (2019). Classification of Food Nutrients Composition using Deep Learning. 4. 66 - 83.
https://www.researchgate.net/publication/338935647_Classification_of_Food_Nutrients_Composition_using_Deep_Learning

[19] Goyal, Sumit. "Predicting properties of cereals using artificial neural networks: A review." (2015).
https://www.semanticscholar.org/paper/Predicting-properties-of-cereals-using-artificial-A-Goyal/c7a136a56042c0fcb4e98896d40d1b26ea0aff0c

[20] Ricardo Lemos Monteiro, Jaqueline Oliveira de Moraes, Jessica Daiane Domingos, Bruno Augusto Mattar Carciofi, João Borges Laurindo, Evolution of the physicochemical properties of oil-free sweet potato chips during microwave vacuum drying, Innovative Food Science & Emerging Technologies, 10.1016/j.ifset.2020.102317, (102317), (2020).
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-4603.2009.00192.x

[21] Errez, Silvia Soledad Moreno Guti et al. "Sustainable and technological strategies for basic cereal crops in the face of climate change: A literature review." African Journal of Agricultural Research 13 (2018): 220-227.

https://www.semanticscholar.org/paper/Sustainable-and-technological-strategies-for-basic-errez-Palacios/321f2a92158efae791b786311c3a79a665c4b436

[22] Patrício, Diego Inácio and Rafael Rieder. "Computer vision and artificial intelligence in

precision agriculture for grain crops: A systematic review." Comput. Electron. Agric. 153 (2018): 69-81.

https://www.semanticscholar.org/paper/Computer-vision-and-artificial-intelligence-in-for-Patr%C3%ADcio-Rieder/2243f46e171cf184412b4b93325dc96e445252ca

[23]Ganesh-Kumar, A. & Mehta, Rajesh & Pullabhotla, Hemant & Prasad, Sanjay & Ganguly, Kavery & Gulati, Ashok. (2012). Demand and supply of cereals in India: 2010-2025.

https://www.researchgate.net/publication/254416821_Demand_and_supply_of_cereals_in_India_2010-2025/citation/download

[24] Choudhury, Nalinikanta & Saurav, Sumeet & Kumar, Rajeev & Budhlakoti, Neeraj. (2017). Modelling and Forecasting of Total Area, Irrigated Area, Production and Productivity of Important Cereal Crops in India towards Food Security. International Journal of Current Microbiology and Applied Sciences. 6. 2591-2600. 10.20546/ijcmas.2017.610.305.

https://www.researchgate.net/publication/320575266_Modelling_and_Forecasting_of_Total_Area_Irrigated_Area_Production_and_Productivity_of_Important_Cereal_Crops_in_India_towards_Food_Security/citation/download

[25] Ch, Dr & Jamshaid, Muhammad & Ali, Asad & Rashid, Qaisar. (2016). forecasting of wheat production: A comparative study of Pakistan and India. International journal of advance research. 4. 698-709. 10.21474/IJAR01/2449.

https://www.researchgate.net/publication/312027794_forecasting_of_wheat_production_A_comparative_study_of_Pakistan_and_India/citation/download

[26] Balaghi, Riad & Jlibene, Mohammed & Tychon, Bernard & Eerens, Herman. (2013). Agrometeorological Cereal Yield Forecasting in Morocco. 10.13140/RG.2.1.3645.6805.

https://www.researchgate.net/publication/258807169_Agrometeorological_Cereal_Yield_Forecasting_in_Morocco/citation/download

[27] Nobile, M.A. & Buonocore, G.G. & Limbo, Sara & Fava, Patrizia. (2003). Shelf Life Prediction of Cereal-based Dry Foods Packed in Moisture-sensitive Films. Journal of Food Science. 68. 1292 - 1300. 10.1111/j.1365-2621.2003.tb09642.x.

https://www.researchgate.net/publication/227844034_Shelf_Life_Prediction_of_Cereal-based_Dry_Foods_Packed_in_Moisture-sensitive_Films/citation/download

[28] Hernández Córdova, Naivy & Soto, Francisco & Florido, René & Llerena, Rodolfo Raimundo & Núñez, Alberto & Maqueira López, Lazaro & Lazo, Greco & Ceijas, Teresa & Durruty, Yoima & López, Aymara & Solano, Oscar & Montenegro, Ranses & Gómez, Lázara & Causse, Antonio. (2016). Use of a simulation model for predicting the behavior of some cereal crops under the conditions of Cuba. 10.13140/RG.2.1.2193.8806.

https://www.researchgate.net/publication/304381182_Use_of_a_simulation_model_for_predicting_the_behavior_of_some_cereal_crops_under_the_conditions_of_Cuba/citation/download

[29] Monnet, Anne-Flore & Laleg, Karima & Michon, Camille & Micard, Valérie. (2019). Legume enriched cereal products: A generic approach derived from material science to predict their structuring by the process and their final properties. Trends in Food Science & Technology. 86. 10.1016/j.tifs.2019.02.027

https://www.researchgate.net/publication/331001892_Legume_enriched_cereal_products_A_generic_approach_derived_from_material_science_to_predict_their_structuring_by_the_process_and_their_final_properties/citation/download

[30] Mkumbwa, Solomon. (2011). Cereal food commodities in Eastern Africa: consumption -

production gap trends and projections for 2020. 10.13140/RG.2.2.32277.37609.

https://www.researchgate.net/publication/327750834_Cereal_food_commodities_in_Eastern_Africa_consumption_-_production_gap_trends_and_projections_for_2020/citation/download

[31] Nasiru, Suleman & Sarpong, Solomon. (2012). Production and Consumption of Corn in Ghana: Forecasting Using ARIMA Models. Asian Journal of Agricultural Sciences. 4.

https://www.researchgate.net/publication/266507667_Production_and_Consumption_of_Corn_in_Ghana_Forecasting_Using_ARIMA_Models/citation/download

[32] Asadallah Najafi (2013).Wheat production price performance prediction in the Iranian north province. 10.5897/AJAR11.429.

https://academicjournals.org/journal/AJAR/article-full-text-pdf/E0677D942536

[33] Anna Luiza Bizerra Brito, Lívia Rodrigues Brito, Fernanda Araújo Honorato, Márcio José Coelho Pontes, Liliana Fátima Bezerra Lira Pontes.Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis,Food Research International,Volume 51, Issue 2,2013,

http://www.sciencedirect.com/science/article/pii/S0963996913001075

[34]Matthew P. Reynolds, Emma Quilligan, Pramod K. Aggarwal, Kailash C. Bansal, Anthony J. Cavalieri, Scott C. Chapman, Saharah M. Chapotin, Swapan K. Datta, Etienne Duveiller, Kulvinder S. Gill, Krishna S.V. Jagadish.An integrated approach to maintaining cereal productivity under climate change.2016.

http://www.sciencedirect.com/science/article/pii/S2211912415300171

[35]Hollaway, G. & Vanstone, V. & Nobbs, J. & Smith, J. & Brown, J.. (2008). Pathogenic nematodes of cereal crops in south-west Victoria, Australia. Australasian Plant Pathology - AUSTRALAS PLANT PATHOL. 37. 505-510. 10.1071/AP08048.

https://www.researchgate.net/publication/248894044_Pathogenic_nematodes_of_cereal_crops_in_south-west_Victoria_Australia

[36]Pagani, Valentina & Guarneri, Tommaso & Fumagalli, Davide & Movedi, Ermes & Testi, Luca & Klein, Tommy & Calanca, Pierluigi & Villalobos, Francisco & López-Bernal, Álvaro & Niemeyer, Stefan & Bellocchi, Gianni & Confalonieri, Roberto. (2017). Improving cereal yield forecasts in Europe – The impact of weather extremes. European Journal of Agronomy. in press. 10.1016/j.eja.2017.06.010.

https://www.researchgate.net/publication/318130451_Improving_cereal_yield_forecasts_in_Europe_-_The_impact_of_weather_extremes

[37] Wang, Yanjing & Wang, Zhihong & Wang, Shuai & Li, Jianmin & Li, Xiangdong & Zang, Hecang & Fang, Baoting. (2020). The changes of wheat seedlings in drought condition by exogenous coronatine (COR). Acta Agriculturae Scandinavica, Section B - Soil & Plant Science. 1-7. 10.1080/09064710.2020.1767191.

https://www.researchgate.net/publication/341887547_The_changes_of_wheat_seedlings_in_drought_condition_by_exogenous_coronatine_COR

[38] Mutlu, Ayse C. et al. “Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks.” European Food Research and Technology 233 (2011): 267-274.

https://www.semanticscholar.org/paper/Prediction-of-wheat-quality-parameters-using-and-Mutlu-Boyaci/3bbcc5d7888607959d60cb3f6c3fa21393156aa4

[39] Paliwal, Jitendra et al. “Cereal grain and dockage identification using machine vision.” (2012).

https://www.semanticscholar.org/paper/Cereal-grain-and-dockage-identification-using-Paliwal-Visen/25bea14d6dbd9e4630a7324c4dbbbbe225107c84

[40] Douik, Ali and Mehrez Abdellaoui. “Cereal Grain Classification by Optimal Features and Intelligent Classifiers.” Int. J. Comput. Commun. Control 5 (2010): 506-516.

https://www.semanticscholar.org/paper/Cereal-Grain-Classification-by-Optimal-Features-and-Douik-Abdellaoui/3f88e5780d73de058e3ef0ea37546e724cccc8d4

[41] Ba Tuan Le,Application of deep learning and near infrared spectroscopy in cereal analysis,Vibrational Spectroscopy,Volume 106,2020,103009,ISSN 0924-2031.

https://linkinghub.elsevier.com/retrieve/pii/S0924203119302565

[42] Wu, Lan Lan et al. “Classification of Single Cereal Grain Kernel Using Shape Parameters Based on Machine Vision.” Advanced Materials Research 605-607 (2012): 2179 – 2182

https://www.semanticscholar.org/paper/Classification-of-Single-Cereal-Grain-Kernel-Using-Wu-Wu/ef34319968a5a995320efce3d622b4ba24595d50

[43] R. Paul Gorman, T. J. Sejinowski, Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets, Neural Networks, Vol 1, No 1, 1988, 75-90

https://papers.cnl.salk.edu/PDFs/Analysis%20of%20Hidden%20Units%20in%20a%20Layered%20Network%20Trained%20to%20Classify%20Sonar%20Targets%201988-2996.pdf

[44] Kim, Y., Roh, J. H., & Kim, H. Y. (2017). Early forecasting of rice blast disease using long short-term memory recurrent neural networks. *Sustainability (Switzerland)*, *10*(1), [34].

https://doi.org/10.3390/su10010034

[45] Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 1990, pp. 1-6 vol.1.doi: 10.1109/IJCNN.1990.137535.

https://ieeexplore.ieee.org/document/5726498

[46] Monnet, Anne-Flore & Laleg, Karima & Michon, Camille & Micard, Valérie. (2019). Legume enriched cereal products: A generic approach derived from material science to predict their structuring by the process and their final properties. Trends in Food Science & Technology. 86. 10.1016/j.tifs.2019.02.027.

http://www.ijcset.com/docs/IJCSET14-05-03-167.pdf

[47] Silva Fuzzo, D.F., Carlson, T.N., Kourgialas, N.N. et al. Coupling remote sensing with a water balance model for soybean yield predictions over large areas. Earth Sci Inform 13, 345–359 (2020).

https://link.springer.com/article/10.1007/s12145-019-00424-w

[48] Franzen, Dave & Kitchen, Newell & Holland, Kyle & Schepers, J.s & Raun, William. (2016). Algorithms for In-Season Nutrient Management in Cereals. Agronomy Journal. 108. 10.2134/agronj2016.01.0041.

https://www.researchgate.net/publication/305340116_Algorithms_for_InSeason_Nutrient_Management_in_Cereals

[49] Karimi, Hadi & Skovsen, Søren & Dyrmann, Mads & Jørgensen, Rasmus. (2018). A Novel Locating System for Cereal Plant Stem Emerging Points’ Detection Using a

Convolutional Neural Network. Sensors. 18. 1611. 10.3390/s18051611.

https://www.researchgate.net/publication/325267210_A_Novel_Locating_System_for_Cereal_Plant_Stem_Emerging_Points'_Detection_Using_a_Convolutional_Neural_Network

[50] Kantanantha, Nantachai & Serban, N. & Griffin, Paul. (2010). Yield and Price Forecasting for Stochastic Crop Decision Planning. Journal of Agricultural, Biological, and Environmental Statistics. 15. 362-380. 10.1007/s13253-010-0025-7.

https://www.researchgate.net/publication/225379205_Yield_and_Price_Forecasting_for_Stochastic_Crop_Decision_Planning