Application of feature engineering and artificial intelligence in industrial systems programs [1]

Dr Lalamani Budeli

Abstract

Machine learning fits mathematical models to data to derive insights or make predictions. A feature is a numeric representation of an aspect of raw data and features sit between data and models in the machine learning pipeline. In the 21st century, the use of software systems to improve the performance of engineering systems intelligence has realized major improvements in industry. In the modern industrial engineering system, the compatibility of equipment to communicate with other equipment using standard industrial protocol and sharing data is a requirement for manufacturers throughout the world. The improving need for industrial machine intelligence requires machine learning algorithms that use input data that gets processed to create outputs. The level of intelligence required from engineering systems require incorporation of data science process which are agile, following iterative methodology to deliver predictive analytics solutions and intelligent applications efficiently.

Engineering process control and instrumentation equipments, networks, controllers, and databases produce a lot of data that requires learning algorithms to improve machine intelligence through feature engineering and its role in enhancing data in machine learning. Engineered features that enhance training provide information that better differentiates the patterns in the data required domain expertise for sound and productive decisions. The purpose of this paper is to demonstrate how the application of feature engineering and artificial intelligence can help companies fully realize the benefits of their analytics program.

Keywords: Feature engineering, industrial systems, project and programs

Introduction

In the 20th century, the process and system approach gained momentum in both science and engineering. A lot of systems and process models were designed to resolve industrial problems. Systems thinking focuses on how the whole system works together while a process is a smaller part of the larger system. A system maintains its existence through the mutual interaction of its parts with a sequence of activities, to keep the coherence of the parts and ensure the required functions. A process on the other hand, is a sequence of activities intended to produce a particular result. This means that inside the system, there are processes. Engineering processes and systems produce a lot of data that if it is interpreted accurately can assist the industry to improve efficiency, resolve problems quickly and save money. Part of the complexity of

[1] How to cite this paper: Budeli, L. (2021). Application of feature engineering and artificial intelligence in industrial systems programs, *PM World Journal*, Vol. X, Issue II, February.

engineered systems research derives from integrating factors that represent their use in products or services factoring in their impacts on manufacturing efficiency, the environment, society, or the human body, as appropriate to the system chosen. Therefore, consideration of the complex interactions and effects of the system's operation is factored into the development, integration, and ongoing management of the components. Engineering process and systems produce a lot of data and the figure below indicates the path between data and answers.

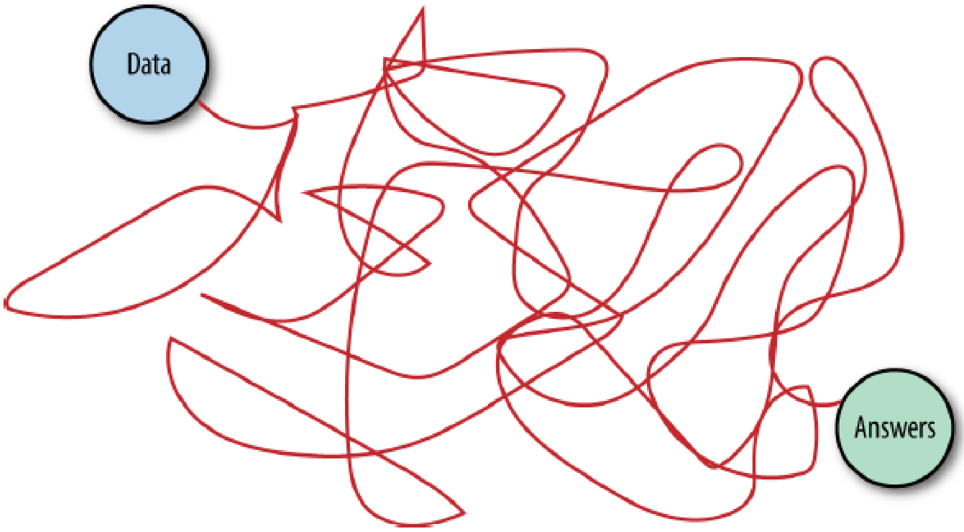

Figure 1 Path Between Data and Answers (Zheng & Casari ,2018:4)

As an emergent technological advancement, machine learning algorithms use some input data to produce results. In most systems, the data available might not be enough for designing a good machine learning model which requires feature engineering. Electrical and computer engineers work at the forefront of technological innovation, contributing to the design, development, testing, and manufacturing processes for new generations of devices and equipment. The process of creating new features from raw data is important to increase the predictive power of the learning algorithm. Engineered features should capture additional information that is not easily apparent in the original feature set. The process of selecting the key subset of features to reduce the dimensionality of the training problem.

The research addressed three hypotheses

- H1: Feature engineering techniques are a compulsory concept for machine learning algorithms in engineering projects and programs.
- H2: Featuring engineering will make things easy for artificial intelligence solutions in engineering, construction, and logistics.

Research method

According to Kumar, Boehm and Yang (2017:1722), research in systems engineering has to build on available scientific methods, both technical and social science. The elementary research methods are based on systems engineering research and how the logical order of steps to define a research project in systems engineering can be implemented. The starting point is a need for improvement in the field, triggered by an industrial problem. Researchers reformulate the problem into an industrial goal. The figure below indicates the research method.

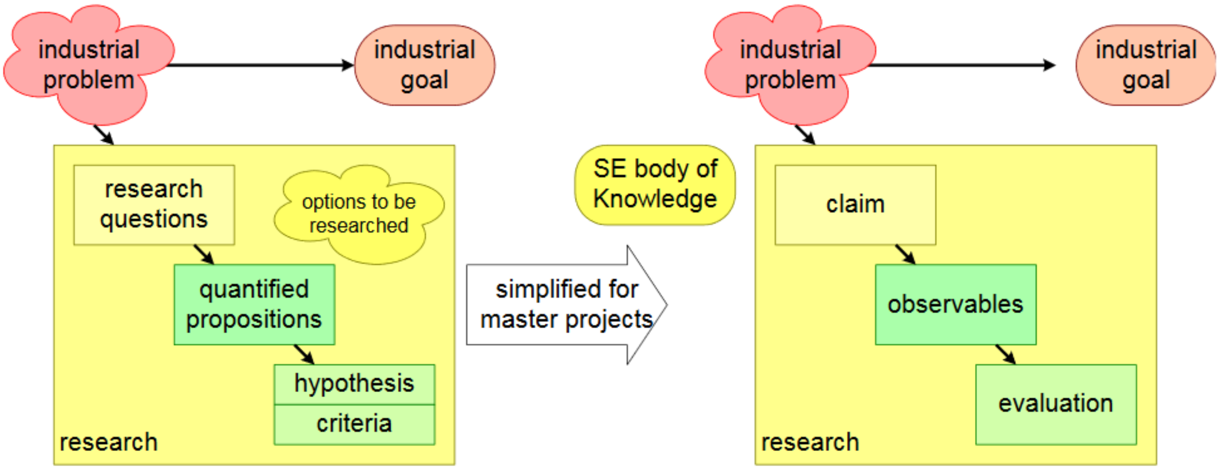

Figure 2 Industrial problem to validated research-Empirical research method (Muller , 2013:1094)

The figure below is a context diagram of the research and many entities that are relevant for the research. To study how effective a concept selection technique is, we have to observe and describe the people using them, the process in which the concept selection is applied, the stakeholders and their concerns and objectives, the artifacts used, and the concepts themselves.

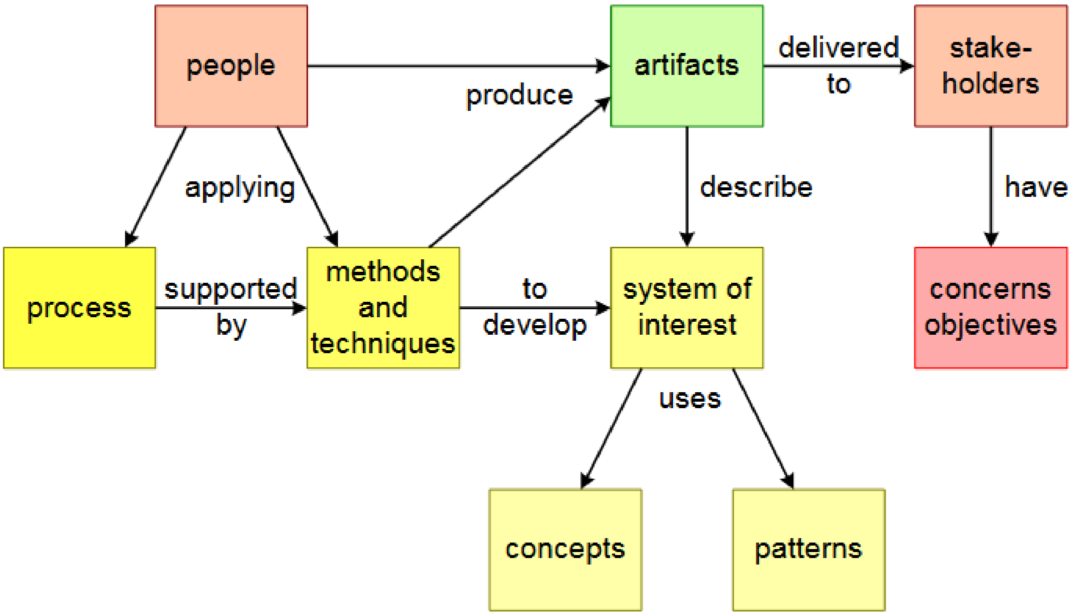

Figure 3 Context diagram with entities relevant for the research (Muller , 2013:1093)

Background of the study

Machine learning and electrical engineering professionals leverage artificial intelligence (AI) to build and optimize systems and provide AI technology with new data inputs for interpretation. Engineers build systems of connected sensors and cameras that ensure that an autonomous vehicle's AI sees the environment. They also ensure that the information is communicated from these onboard sensors at lightning speed, as any delay in processing could result in a severe accident. Below are feature engineering techniques in machine learning.

Feature engineering

According to Nargesian, Samulowitz, Khurana, Khalil and Turaga (2017:2532), the best way to achieve expertise in feature engineering is practicing different techniques on various datasets and observing their effect on model performances. The following are tools and techniques:

Imputation - According to Zheng and Casari (2018:11), the two types of imputation are numerical and categorical. Fan, Sun, Zhao, Song and Wang (2019:36) said that numerical imputation provides some machine learning platforms automatically drop the rows which include missing values in the model training phase and it decreases the model performance because of the reduced training size. On the other hand, most of the algorithms do not accept datasets with missing values and gives an error. Missing values are one of the most common problems you can encounter when you try to prepare your data for machine learning. Nargesian, Samulowitz, Khurana, Khalil and Turaga (2017:2532) said that the reason for the missing values might be human errors, interruptions in the data flow, privacy concerns, and so on. Whatever is the reason, missing values affect the performance of the machine learning models.

Handling Outliers - Banerjee, Chattopadhyay, Pal and Garain (2018:27) said that before mentioning how outliers can be handled, I want to state that the best way to detect the outliers is to demonstrate the data visually. All other statistical methodologies are open to making mistakes, whereas visualizing the outliers gives a chance to decide with high precision. According to Tymoshenko, Bonadiman and Moschitti (2016:1271), Statistical methodologies are less precise, but on the other hand, they have a superiority, they are fast. There are two different ways of handling outliers. These will detect them using standard deviation, and percentiles.

Outlier Detection with Standard Deviation - If a value has a distance to the average higher than x * standard deviation, it can be assumed as an outlier.

Outlier Detection with Percentiles - Another mathematical method to detect outliers is to use percentiles. You can assume a certain percentage of the value from the top or the bottom as an outlier.

According to Zhou and Li (2020:58), data in computing is information that has been translated into a form that is efficient for movement or processing, which are also observations of real-world phenomena. For example, the energy industrial process produces a lot of data that when interpreted becomes useful information. Mathematical modeling in particular statistical modelling are used to understand the world through cleaning and transforming data through features which are numeric representation of raw data.

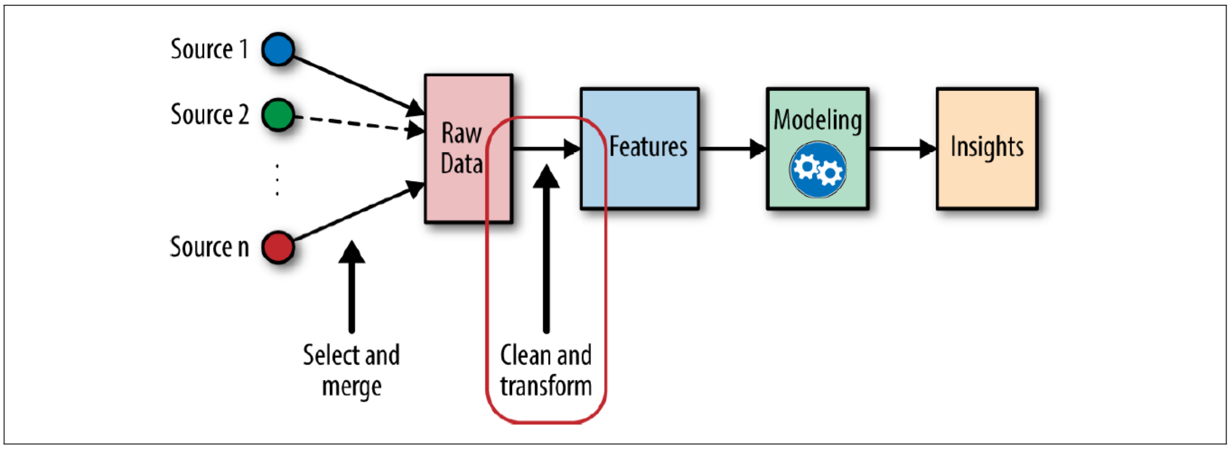

Figure 4 The place of feature engineering in the machine learning workflow (Zheng & Casari , 2018:4)

Binning - RM, Maddikunta, Parimala, Koppu, Reddy, Chowdhary and Alazab (2020:16) indicated that the trade-off between performance and overfitting is the key point of the binning process. In my opinion, for numerical columns, except for some obvious overfitting cases, binning might be redundant for some kind of algorithms, due to its effect on model performance. Zhou and Li

(2020:58) said that the main motivation of binning is to make the model more robust and prevent overfitting, however, it has a cost to the performance. Every time you bin something, you sacrifice information and make your data more regularized

Log Transform - Su, Tao, Zhang, Cheng, Ma and Wang (2020:871) directed that the Logarithm transformation (or log transform) is one of the most commonly used mathematical transformations in feature engineering. Mohamad, Ahmad, Jawawi and Hashim (2020:885) said that the benefits of log transform is that it helps to handle skewed data and after transformation, the distribution becomes more approximate to normal.

One-hot encoding - Lepadat (2019:24) assumed that one-hot encoding is one of the most common encoding methods in machine learning. This method spreads the values in a column to multiple flag columns and assigns 0 or 1 to them. These binary values express the relationship between grouped and encoded column.

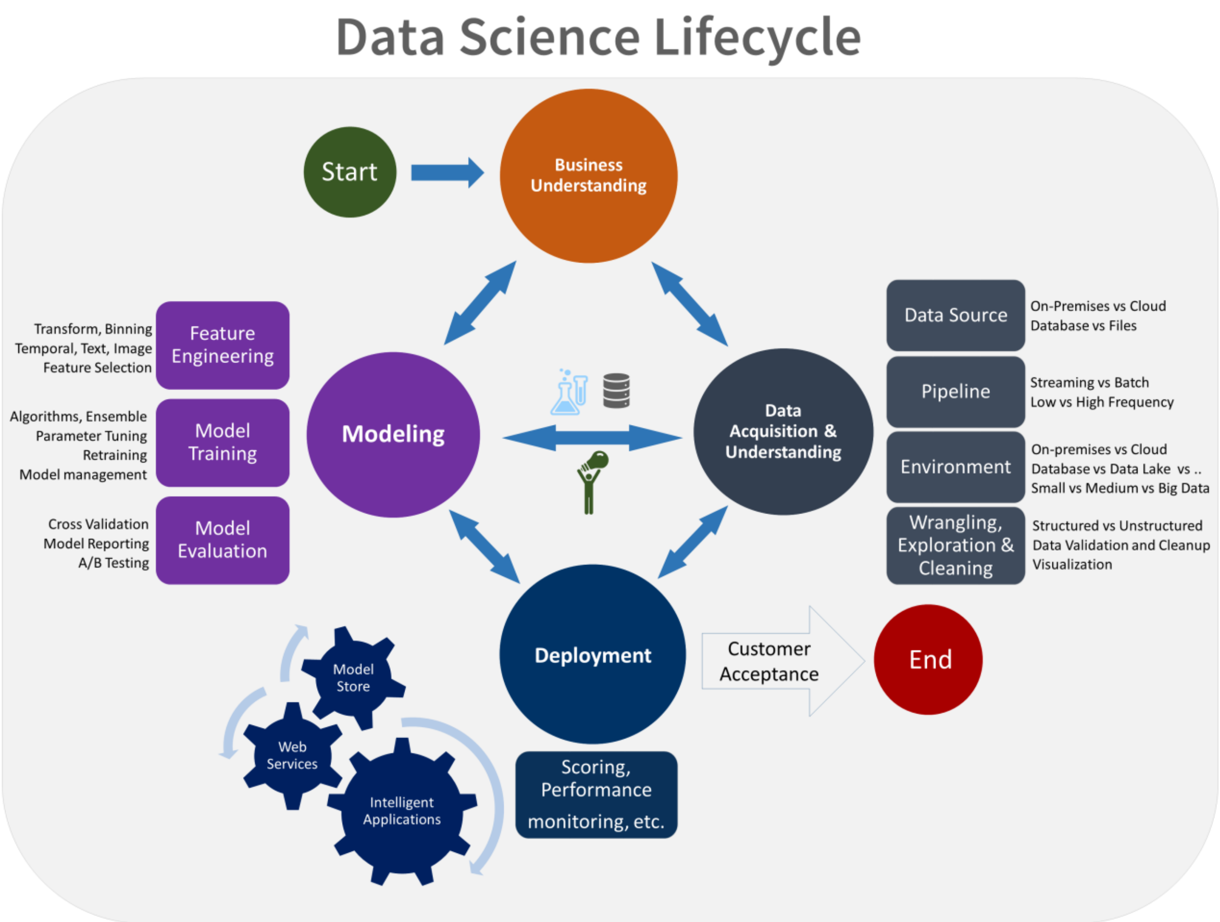

Figure 5 Data science life cycle

According to Koitka and Friedrich (2016:335), artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. Gao, Liu, Shen and Li (2020:22) defined Artificial Intelligence as the development of computer systems that can perform tasks that would require human intelligence. Zhou and Li (2020:58) assumed that the growth in computer architecture is so fundamental that it is dramatically reshaping relationships among people and organizations providing a foundation for understanding and learning of intelligent behavior in living and engineered systems.

Data collection and analysis

Researchers observe or measure and create artifacts as part of their research process. One of the challenges is to collect data in such a way that researchers can use it for analysis. There is a spectrum of possibilities on how the researcher can collect data, can extract data from other people, and what artifacts the researcher can produce. In the industrial process, most data can be collected through information systems components which include hardware, software, database, network, people and these five components integrated to perform (i.e. Input, processing, output, Feedback). Software consist of various programs and procedures whereas hardware consists of I/O devices, operating system and the other media devices. Database consists of data organized in a structured way. Network consist of hubs, communication media, and network devices.

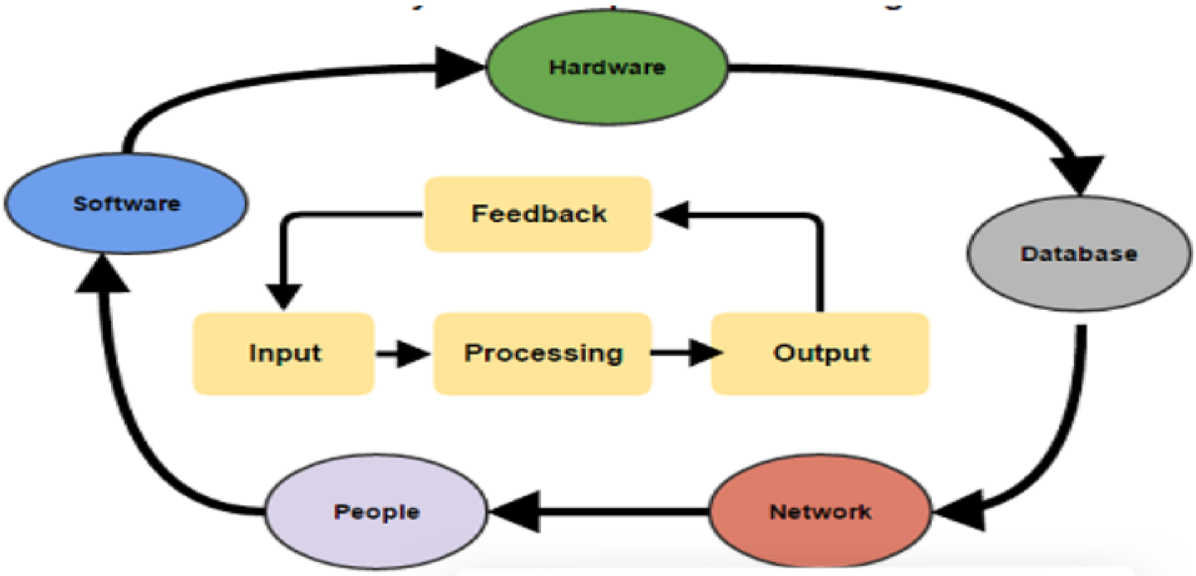

Figure 6 Integrated information system components for Data collection (Muller ,2013:7)

During the input, feeding instructions are fed to the computer and during the processing stage worked upon by software programs and other queries. While in the output phase data is represented in a structured form of documents and the same in the reports.

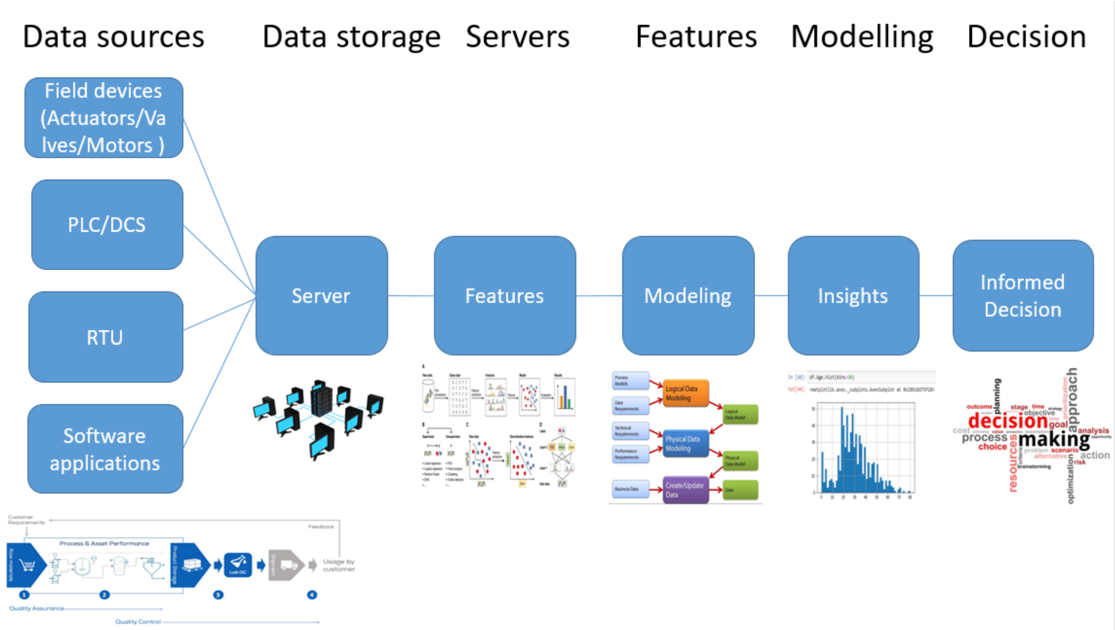

Figure 7 Industrial data collection for feature development during the machine learning process

Develop data features

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data. It is the process of creating features also called attributes that don't already exist in the dataset. This means that if the dataset already contains enough useful features, you don't necessarily need to engineer additional features. A useful feature is a feature that the machine learning model can learn from to more accurately predict the value of the target variable. The following steps are followed to develop features

Step 1: Feature Importance

Feature importance refers to a class of techniques for assigning scores to input features to a predictive model that indicates the relative importance of each feature when making a prediction. Its scores can be calculated for problems that involve predicting a numerical value, called regression, and those problems that involve predicting a class label, called classification. Feature importance scores will provide:

- Insight into the dataset
- Insight into the model.
- Improve a predictive model

Step 2: Preparation

Confirming the environment and preparing some test datasets is important to be used as the basis for demonstrating and exploring feature importance scores. Each test problem has five important and five unimportant features, and it may be interesting to see which methods are consistent at finding or differentiating the features based on their importance or statistical limits.

Step 3: Coefficients as Feature Importance

Linear machine learning algorithms fit a model where the prediction is the weighted sum of the input values. Examples include linear regression, logistic regression, and extensions that add regularization.

```
# linear regression feature importance
from sklearn.datasets import make_regression
from sklearn.linear_model import LinearRegression
from matplotlib import pyplot
# define dataset
X, y = make_regression(n_samples=1000, n_features=10, n_informative=5, random_stat
# define the model
model = LinearRegression()
# fit the model
model.fit(X, y)
# get importance
importance = model.coef_
# summarize feature importance
for i,v in enumerate(importance):
    print('Feature: %0d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()
```

Figure 8 Example of linear regression coefficients for feature importance

This is a classification problem with classes 0 and 1. Notice that the coefficients are both positive and negative. The positive scores indicate a feature that predicts class 1, whereas the negative scores indicate a feature that predicts class 0.

```
Feature: 0, Score: 0.16320
Feature: 1, Score: -0.64301
Feature: 2, Score: 0.48497
Feature: 3, Score: -0.46190
Feature: 4, Score: 0.18432
Feature: 5, Score: -0.11978
Feature: 6, Score: -0.40602
Feature: 7, Score: 0.03772
Feature: 8, Score: -0.51785
Feature: 9, Score: 0.26540
```

Figure 9 Important and unimportant features

A bar chart is then created for the feature importance scores.

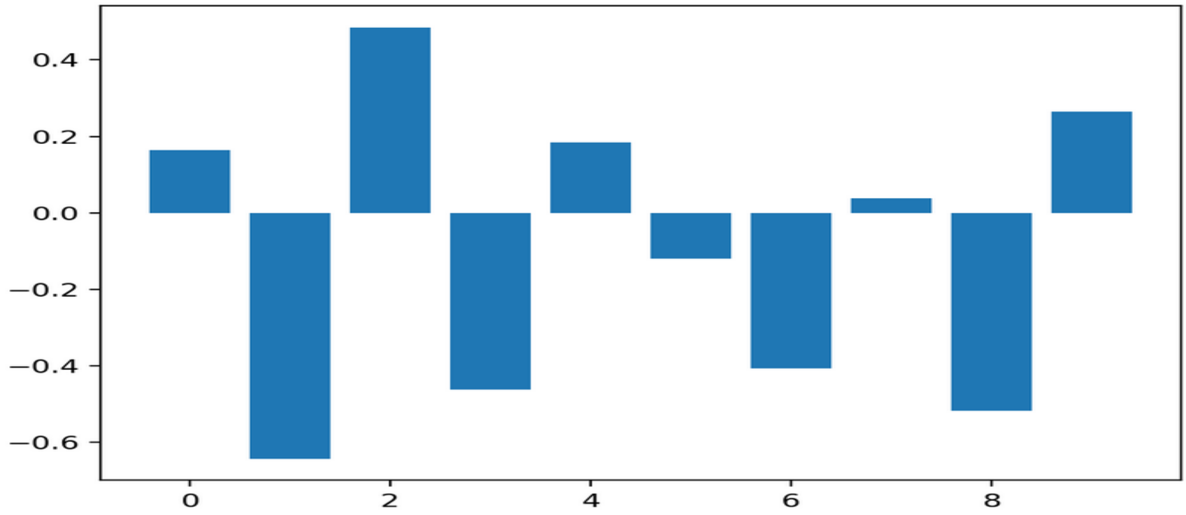

Figure 10 Process A Regression Coefficients as Feature Importance Scores

Step 4: Decision Tree Feature Importance

Decision tree algorithms like classification and regression trees (CART) offer importance scores based on the reduction in the criterion used to select split points. This same approach can be used for ensembles of decision trees, such as the random forest and stochastic gradient boosting algorithms. The results suggest perhaps four of the 10 features as being important to prediction.

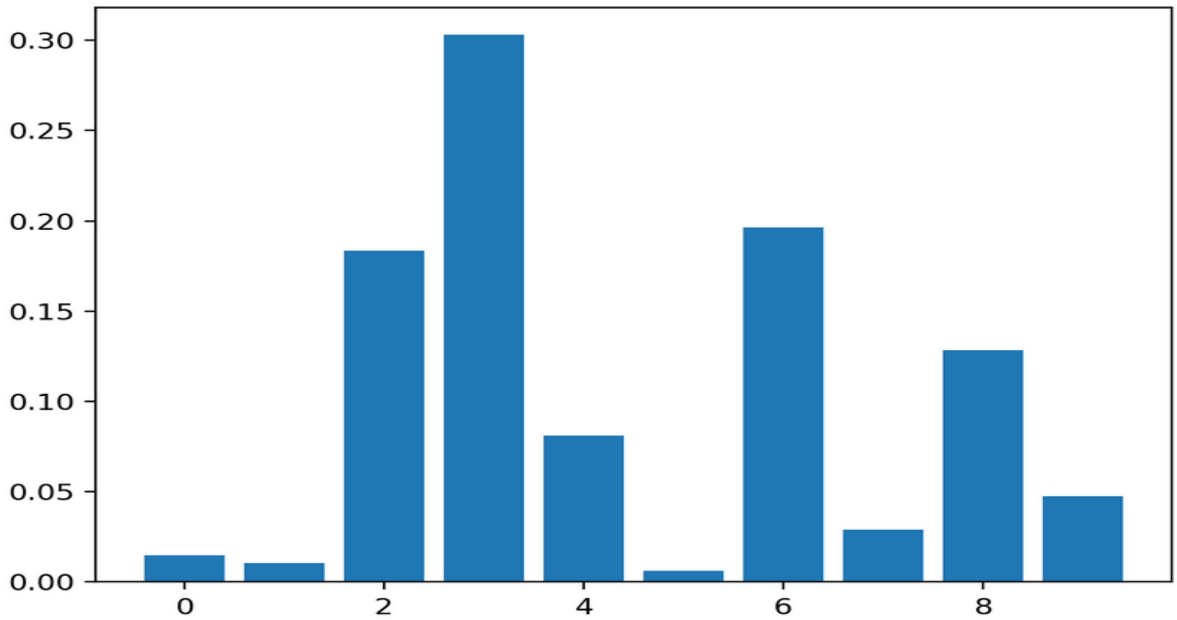

Figure 11 DecisionTree Classifier Feature Importance Scores

Step 5: Permutation Feature Importance

Permutation feature importance is a technique for calculating relative importance scores that is independent of the model used. First, a model is fit on the dataset, such as a model that does not support native feature importance scores. Then the model is used to make predictions on a dataset, although the values of a feature (column) in the dataset are scrambled. This is repeated for each feature in the dataset. Then this whole process is repeated 3, 5, 10, or more times. The result is a mean importance score for each input feature

Step 6: Feature Selection with Importance

Feature importance scores can be used to help interpret the data, but they can also be used directly to help rank and select features that are most useful to a predictive model. First, we can split the training dataset into train and test sets and train a model on the training dataset, make predictions on the test set and evaluate the result using classification accuracy. We will use a logistic regression model as the predictive model. This provides a baseline for comparison when we remove some features using feature importance scores.

Following these steps, it can be seen that the model achieves the same performance on the dataset, although with half the number of input features. As expected, the feature importance scores calculated by random forest allowed us to accurately rank the input features and delete those that were not relevant to the target variable.

Application of AI

Artificial intelligence use in the engineering sector utilize both software and hardware components. As machines become more sophisticated, they can support smart production lines and complex manufacturing tasks. They will also be able to design and improve tasks over time with little or no human intervention through machine learning. Below are some possible area in engineering and project management, where AI will be more useful:

Prevent cost overruns

According to McKinsey (2017-2), most megaprojects go over budget despite employing the best project teams. Artificial Neural Networks are used on projects to predict cost overruns based on factors such as project size, contract type, and the competence level of project managers. Zheng and Casari (2018:11) designated that historical data such as planned start and end dates are used by predictive models to envision realistic timelines for future projects. Xia and Lo (2018:348) held aI helps staff remotely access real-life training material which helps them enhance their skills and knowledge quickly which will reduce the time taken to onboard new resources onto projects. As a result, project delivery is expedited.

AI for Better Design of Buildings Through Generative Design

Kumar, Boehm and Yang (2017:1721) indicated that, building Information Modeling is a 3D model-based process that gives architecture, engineering and construction professionals insights to efficiently plan, design, construct and manage buildings and infrastructure. Koitka, and Friedrich (2016:335) whispered that to plan and design the construction of a building, the 3D models need to take into consideration the architecture, engineering, mechanical, electrical, and plumbing (MEP) plans and the sequence of activities of the respective teams. Lucas, Portier, Laporte, He-Guelton, Caelen, Granitzer and Calabretto (2020:395) specified that the challenge is to ensure that the different models from the sub-teams do not clash with each other. Kalidindi (2020:131) reviled that the industry is trying to use machine learning in the form of generative design to identify and mitigate clashes between the different models generated by the different teams in the planning and design phase to prevent rework. There is software that uses machine learning algorithms to explore all the variations of a solution and generates design alternatives. Banerjee, Chattopadhyay, Pal and Garain (2018:29) believed that it leverages machine learning to specifically create 3D models of mechanical, electrical, and plumbing systems while simultaneously making sure that the entire routes for MEP systems do not clash with the building architecture while it learns from each iteration to come up with an optimal solution.

Risk Mitigation

Xia and Lo (2018:348) direct that every construction project has some risk that comes in many forms such as Quality, Safety, Time, and Cost Risk. Kalidindi (2020:131) supposed that the larger the project, the more risk, as multiple sub-contractors are working on different trades in parallel on job sites. Tymoshenko, Bonadiman and Moschitti (2016:1273) understood that there are AI and machine learning solutions today that general contractors use to monitor and prioritize risk on the job site, so the project team can focus their limited time and resources on the biggest risk factors. Mohamad, Ahmad, Jawawi and Hashim (2020:887) agreed that AI can be used to automatically assign priority to issues. Subcontractors are rated based on a risk score so construction managers can work closely with high-risk teams to mitigate risk.

Project Planning

Lucas, Portier, Laporte, He-Guelton, Caelen, Granitzer and Calabretto (2020:395) indicated that an AI Startup launched in 2018 with the promise that its robots and artificial intelligence hold the key to solving late and over-budget construction projects. The company uses robots to autonomously capture 3D scans of construction sites and then feeds that data into a deep neural network that classifies how far along different sub-projects are. If things seem off track, the management team can step in to deal with small problems before they become major issues. Kalidindi (2020:132) supposed that algorithms of the future will use an AI technique known as "reinforcement learning." This technique allows algorithms to learn based on trial and error. Kumar, Boehm and Yang (2017:1723) thought that it can assess endless combinations and alternatives based on similar projects. It aids in project planning since it optimizes the best path and corrects itself over time.

AI Will Make Jobsites More Productive

Kumar, Boehm and Yang (2017:1718) alleged that there are companies that are starting to offer self-driving construction machinery to perform repetitive tasks more efficiently than their human counterparts, such as pouring concrete, bricklaying, welding, and demolition. Excavation and prep work is being performed by autonomous or semi-autonomous bulldozers, which can prepare a job site with the help of a human programmer to exact specifications. Koitka, and Friedrich (2016:335) agreed that this frees up human workers for the construction work itself and reduces the overall time required to complete the project and project managers can also track job site work in real-time by using facial recognition, onsite cameras, and similar technologies to assess worker productivity and conformance to procedures.

AI for Construction Safety

Nargesian,Samulowitz, Khurana, Khalil and Turaga (2017:2532) said that construction workers are killed on the job five times more often than other laborers. According to OSHA, the leading causes of private sector deaths (excluding highway collisions) in the construction industry were falls, followed by struck by an object, electrocution, and caught-in/between. Zheng and Casari (2018:11) indicated that companies are developing an algorithm that analyzes photos from their job sites, scans them for safety hazards such as workers not wearing protective equipment, and correlates the images with their accident records. Zhou and Li (2020:58) said that companies have the potential to compute risk ratings for projects so safety briefings can be held when an elevated threat is detected.

AI Will Address Labor Shortages

According to Nargesian,Samulowitz, Khurana, Khalil and Turaga (2017:2532) , labor shortage and a desire to boost the industry’s low productivity are compelling construction firms to invest in AI and data science. A 2017 McKinsey report says that construction firms could boost productivity by as much as 50 percent through real-time analysis of data. Construction companies are starting to use AI and machine learning to better plan for the distribution of labor and machinery across jobs. Lucas, Portier, Laporte, He-Guelton, Caelen, Granitzer and Calabretto (2020:395) understood that a robot constantly evaluating job progress and the location of workers and equipment enables project managers to tell instantly which job sites have enough workers and equipment to complete the project on schedule, and which might be falling behind where additional labor could be deployed. Lepadat (2019:24) indicated that experts expect construction robots to become more intelligent and autonomous with AI techniques.

Off-site Construction

Construction companies are increasingly relying on off-site factories staffed by autonomous robots that piece together components of a building, which are then pieced together by human workers on-site. Structures like walls can be completed assembly-line style by autonomous machinery more efficiently than their human counterparts, leaving human workers to finish the detail work like Plumbing, HVAC and Electrical systems when the structure is fitted together.

AI and Big Data in Construction

Xia and Lo (2018:352) assumed that at a time when a massive amount of data is being created every day, AI Systems are exposed to an endless amount of data to learn from and improve every day. Every job site becomes a potential data source for AI. Data generated from images captured from mobile devices, drone videos, security sensors, building information modeling (BIM), and others have become a pool of information. Su, Tao, Zhang, Cheng, Ma and Wang (2020:871) whispered that this presents an opportunity for construction industry professionals and customers to analyze and benefit from the insights generated from the data with the help of AI and machine learning systems.

AI for Post-Construction

Zhou and Li (2020:59) indicated that building managers can use AI long after the construction of a building is complete. Building information modeling, or BIM, stores information about the structure of the building. AI can be used to monitor developing problems and even offers solutions to prevent problems of the value chain for construction material manufacturers and distributors. The construction industries are on the verge of digitalization, which is disrupting traditional processes and also holds many opportunities in store. Artificial intelligence is expected to increase efficiency throughout the entire value chain from the production of building materials to the design, planning and construction phase itself, and facility management as well. But how do you best benefit from AI in your company? Aside from the opportunities for the use of AI along a construction project's lifecycle, AI can also be an enabler for the next big step in the digitalization of construction material producers and/or distributors. It would increase efficiency along the value chain in functions such as procurement, marketing and sales, manufacturing, logistics, customer service and aftersales of building material producers, merchants as well as building companies.

AI for Procurement

Banerjee, Chattopadhyay, Pal and Garain (2018:28) specified that, in procurement, AI can help companies forecast prices for raw materials (sand, gravel, iron, etc.) and other input factors. To do so, AI will analyze data on both past price development and the development of other factors that influence the price of a raw material, or even identify such relationships autonomously. Koitka, and Friedrich (2016:335) showed that AI can then forecast the price and determine the optimal point in time to buy. This can be connected to data on inventory so that AI can include the stock of certain raw materials in its calculations and perform optimized inventory management. Fan, Sun, Zhao, Song and Wang (2019:37) specified that with AI, it is going to be possible to automate more and more additional steps of the purchase-to-pay process. As described above, AI can identify the need for a product and then order said product. Zhou and Li (2020:58) alleged that AI also helps further optimize logistics concerning routes. Artificial intelligence identifies the best means of transportation and also the best routes based on information on times and dates, addresses, traffic, costs, speed, etc. Also, it improves its recommendations with every delivery executed.

Findings

The research findings for each research hypothesis as outlined in the research hypothesis, are analyzed in the following paragraph.

Hypothesis 1

H_0 : Feature engineering techniques are not compulsory concepts for machine learning algorithms in engineering projects and programs.

H_1 Feature engineering techniques are a compulsory concept for machine learning algorithms in engineering projects and programs.

Most machine learning experts ask themselves what is a feature and why we need the engineering of it? All machine learning algorithms use some input data to create outputs. This input data comprise features, which are usually in the form of structured columns. Algorithms require features with some specific characteristics to work properly. The need for feature engineering efforts goals is to prepare the proper input dataset which is compatible with the machine learning algorithm requirements and improving the performance of machine learning models.

From the above results H_1 : Feature engineering techniques are a compulsory concept for machine learning algorithms in engineering projects and programs.

Hypothesis 2

H_0 : Featuring engineering will not make things easy for artificial intelligence solution in engineering, construction and logistics.

H_1 : Featuring engineering will make things easy for artificial intelligence solutions in engineering, construction and logistics.

Engineering complex systems can be identified by what they do and also by how they may or may not be analyzed. Data is like the crude oil of machine learning because it has to be refined into features (predictor variables) to be useful for training a model. Without relevant features, models cannot be trained accurately, no matter how complex the machine learning algorithm. the process of extracting features from a raw dataset is an important part of the machine learning algorithm and a central part of artificial intelligence (AI) applications.

From the above results H_1 : Featuring engineering will make things easy for artificial intelligence solutions in engineering, construction and logistics.

Conclusion

According to a survey in Forbes, data scientists spend 80% of their time on data preparation. This is very impressive to show the importance of feature engineering in data science. Some techniques might work better with some algorithms or datasets, while some of them might be beneficial in all cases. The best way to achieve expertise in feature engineering is by practicing different techniques on various datasets and observing their effect on model performances. The features you use influence more than everything else of the result. No algorithm alone can supplement the information gain given by correct feature engineering.

References

Banerjee, S., Chattopadhyay, T., Pal, A. and Garain, U., 2018. Automation of feature engineering for iot analytics. ACM SIGBED Review, 15(2), pp.24-30.

Benţa, D., Rusu, L., Jecan, S. and Dines, O., 2016, July. Feature of development for risk management in Pavement Maintenance system. In 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 413-417). IEEE

Fan, C., Sun, Y., Zhao, Y., Song, M. and Wang, J., 2019. Deep learning-based feature engineering methods for improved building energy prediction. Applied Energy, 240, pp.35-45.

Gao, H., Liu, C., Shen, Q. and Li, J., 2020. Representative Databases for Feature Engineering and Computational Intelligence in ECG Processing. In Feature Engineering and Computational Intelligence in ECG Monitoring (pp. 13-29). Springer, Singapore.

Kalidindi, S.R., 2020. Feature engineering of material structure for AI-based materials knowledge systems. Journal of Applied Physics, 128(4), p.041103.

Koitka, S. and Friedrich, C.M., 2016, September. Traditional Feature Engineering and Deep Learning Approaches at Medical Classification Task of ImageCLEF 2016. In CLEF (Working Notes) (pp. 304-317).

Kumar, A., Boehm, M. and Yang, J., 2017, May. Data management in machine learning: Challenges, techniques, and systems. In Proceedings of the 2017 ACM International Conference on Management of Data (pp. 1717-1722).

Lepadat, M.A., 2019. Rule-based recommender for feature engineering in big data (Doctoral dissertation, Wien).

Lucas, Y., Portier, P.E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M. and Calabretto, S., 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. Future Generation Computer Systems, 102, pp.393-402.

Mohamad, N., Ahmad, N.B., Jawawi, D.N.A. and Hashim, S.Z.M., 2020, July. Feature Engineering for Predicting MOOC Performance. In IOP Conference Series: Materials Science and Engineering (Vol. 884, No. 1, p. 012070). IOP Publishing.

Muller , G , 2013. Systems Engineering Research Methods , conference on system engineering March 19-22

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E.B. and Turaga, D.S., 2017. August. Learning Feature Engineering for Classification. In IJCAI (pp. 2529-2535).

Su, X., Tao, L., Zhang, T., Cheng, Y., Ma, J. and Wang, C., 2020, September. A data-driven FCE method for UAV condition risk assessment based on feature engineering and variable weight coefficients. In 2020 International Conference on Unmanned Aircraft Systems (ICUAS) (pp. 867-874). IEEE.

RM, S.P., Maddikunta, P.K.R., Parimala, M., Koppu, S., Reddy, T., Chowdhary, C.L. and Alazab, M., 2020. An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. Computer Communications.

Tymoshenko, K., Bonadiman, D. and Moschitti, A., 2016, June. Convolutional neural networks vs. convolution kernels: Feature engineering for answer sentence reranking. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 1268-1278).

Xia, X. and Lo, D., 2018. Feature generation and engineering for software analytics. Feature engineering for machine learning and data analytics, p.335.

Zheng, A. and Casari, A., 2018. Feature engineering for machine learning: principles and techniques for data scientists. " O'Reilly Media, Inc."

Zhou, S. and Li, X., 2020. Feature engineering vs. deep learning for paper section identification: Toward applications in Chinese medical literature. Information Processing & Management, 57(3), p.102206

About the Author

Dr Lalamani Budeli

South Africa

Dr Lalamani Budeli obtained his degree in Electrical Engineering at the Vaal University (VUT), BSc honours in Engineering Technology Management at University of Pretoria (UP), Master in engineering development and Management at North West University (NWU), Master of business administration at Regent Business School (RBS) and a Doctor of Philosophy in Engineering Development and Management at North West University (NWU), Potchefstroom, South Africa. Currently, he is a managing director of BLIT, an engineering, research, and project management company based in South Africa.

His research interests include project portfolio management, agile project management, plant life cycle management, advanced systems analytics, project early warning system, and the use of artificial intelligence in engineering and project management. Currently, he is spending most of the time on research that is looking at the development of system and application that uses the latest technology like blockchain, internet of things (IoT), Big data, and artificial intelligence. Lalamani Budeli can be contacted at Budelil@blit.co.za.