International Journal of Innovative Research in Engineering
Volume 3, Issue 3 (May-June 2022), PP: 59-63
www.theijire.com
ISSN No: 2582-8746

Analogy of Expert Systems in Data Analytics

Dr. D. Bhuvaneswari [1], Dr. V. Shoba [2]
[1,2]Assistant Professor, Department of Computer Science, SRM Arts and Science College, Tamilnadu, India.

How to cite this paper:
Dr.D.Bhuvaneswari [1], Dr.V.Shoba[2], "Analogy of Expert Systems in Data Analytics", IJIRE-V3I03-59-63.

Copyright © 2022 by author(s) and 5th Dimension Research Publication.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract: *Machine learning is frequently used in that work in conjunction with their own intellect. Reinforcement learning algorithms and deep learning neural networks are examples of machine artificial intelligence. The term "machine learning" refers to a system's ability to learn. To getknowledge from problem-solving training data to automate the analytical procedureCreating models and completing related tasks. The term "deep learning" refers to a type of machine learning. Artificial neural network concept networks. Deep learning is beneficial in many scenarios. Deep learning models outperform shallow learning models. Models from the pastand models from the future data analysis procedures All of this is organic. Move beyond the technological and elucidate issues in human-machine interaction Artificial intelligence and interaction identified in the early. Information retrieval systems give usersaccess to collections of hundreds or millions ofdocuments, from which they can retrieve any one by providing a suitable description. Users typically refine their descriptions to meet their needs iteratively, and retrieval systems can use user inputon selected documents to signal the accuracy of thedescription at any stage. The type of description requested from the user and how it is used to search the document database are both influenced by the indexing approach chosen for the collection. From storing keywords with links to particular papers to clustering documents under relatedsubjects, the index can take many different shapes. A lot of information retrieval labour can be digitized.*

Keyword: *Artificial Intelligence, Machine Learningand Deep Learning*

I. INTRODUCTION

Artificial intelligence systems do not need to be pre-programmed; instead, they use algorithms learning algorithms. AI can be classed into three sorts based on its capabilities: Weak AI, General AI, and Strong AI. We now have both weak and universal AI. Strong AI is the AI of the future, with predictions that it will be smarter than humans. Machine learning's purpose is to extract knowledge from data. Machine learning is an artificial intelligence field that allows machines to learn without being explicitly taught from prior data or experiences. Machine learning allows a computer system that would make predictions or judgments based on past data withouthaving to be explicitly written. In order for a machine learning model to produce trustworthy findings or make predictions based on it, it requires a huge amount of structured and semi-structured data. It only works for limited domains; for example, if we develop a machinelearning model to detect dog photographs, it will only return dog picture results; however, if we add new data, such as a cat picture, it will becomeunresponsive.

Machine learning is used in a wide range of applications, including online recommendersystems, Googlesearch engines, email spam filters and face book auto friend tagging suggestions. The three types of learning are supervised learning, reinforcement learning, and unsupervised learning. The document is processed as an index, and query refinement is frequently done by computer, while document classification and index word selection are more often done manually. Manually creating and maintaining document databases, on the other hand is time-consuming, tedious, and prone to mistakes. Algorithms that "mine" documents for indexing information and model user interests to aid in query formulation minimize workload and assure consistency.

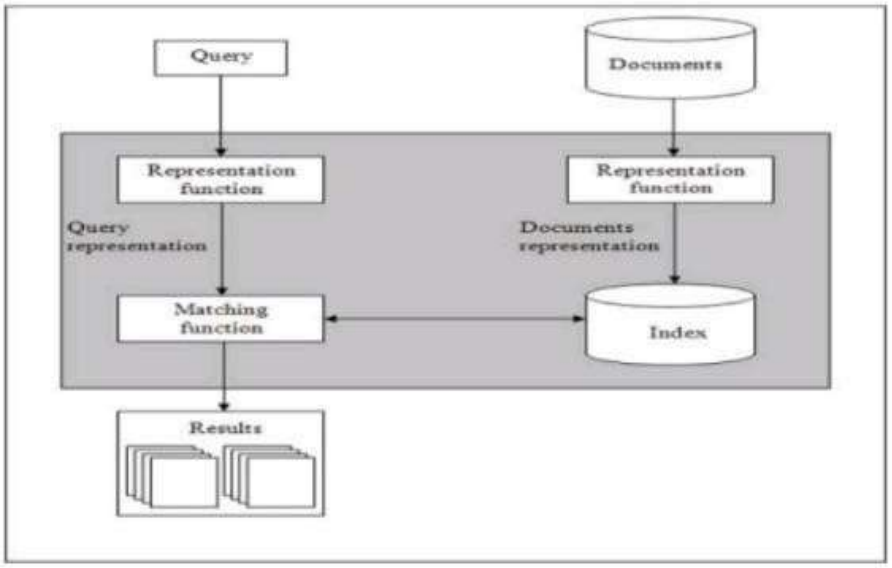
These algorithms are based on machinelearning, a dynamic, developing discipline of computer science that is finding applications in domains ranging from "expert systems," where learning algorithms supplement—or even replace— domain specialists for providing rules and explanations, to "intelligent agents," which learn to perform specific, highly-specialized support roles for individual persons and are considered by some as usheringa new renaissance of artificial intelligence in information technology. Many aspects of information retrieval makeit a good candidate for machine learning.

Classification tasks that are ideally suited to machine learning—in many cases, jobs that had to be completed manually, if at all—are key information retrieval operations. In information retrieval systems instances, traits and values are plentiful therefore learning algorithms rely on them. The profusion of natural language elements that may be retrieved from documents give a wealth of qualities in most such systems, which have thousands of documents.

As a result, there is a great deal of data available; in fact, most systems have much too many attributes and instances for machine learning approaches to successfully process. The rest of this section summarises both the field of information retrieval and the field of machine learning, explaining the major paradigms and showing how they might be used to information retrieval.

Information retrieval steps include, 1 show how machine learning can help with text classification, which is a labor-intensive and time-consuming task when done manually. 2 examines how machine learning can be applied in the query formulation process, such as the automatic creation of user models that provide context for interpreting users' questions and the use of appropriate feedback to enhance query refinement. 3 analyzes document filtering methods in which the user creates a query that is applied to a constantly changing set of documents. The final section summarizes the subject and offers some recommendations.

Fig.1 Information Retrieval Processing



II. INFORMATION RETRIEVAL MODELS

Models for information retrieval predict and explain what a user will find in answer to a query. These are the above-mentioned components of retrieval approach that we described in ad-hoc retrieval and consist of A model for documents, A model for questions, and a matching function that matches queries to documents. On a mathematical level, a retrieval model is made up of the following components. D: Representation of documentationF: The D, Q, and their relationship modeling paradigm R is a query representation language. R isa query representation language. R (q, di): A similarity or ranking function that ranks documents according to their relevance to the query.

2.1 Information Retrieval Model Types

According to their classification, there are three different types of Information documents.

2.1.1 Models utilizing traditional IR-

These would be the simplest and most basic IR models. These are based on instantly recognized and understood mathematical concepts. Traditional IR models include the following: Models come in a variety of forms, including Boolean, vector, andprobabilistic.

2.1.2 Traditional IR Models and Non-Classical IR Models are diametrically opposed. These are founded on, among other things, similarity, probability, and Boolean operations. Non-classical IR models encompass information logic models,situation theory models, and interaction models**.**

2.1.3 Alternative IR Models-This is a development of the classic IR model that includes techniques from other disciplines. The following are some examplesof alternative IR models-Latent Semantic Indexing (LSI) models include cluster models, fuzzy models,and Latent Semantic Indexing (LSI) models, among other things.

2.2 Machine Learning Models

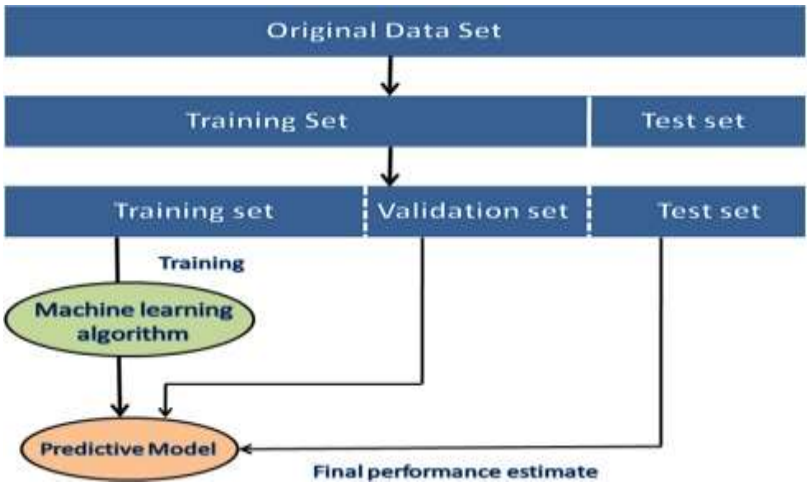
The goal of the IR initiative was to reduce theamount of time it takes to screen articles by narrowing down the literature to a small number of papers that provide reliable information, in this case on biomarkers of exposure. In this instance,the IR problem can be described as a classification task in which we must determine whether or not a sentence is significant. A dataset is a collection of data that has been logically organized. Any type of data, from a series of arrays to a database table, can be stored in a dataset. A representative dataset was seen in the table below. A tabular dataset is a database table or matrix in which each row in a table the dataset's fields and each column represents a variable. "Comma Separated File," or CSV, is the most prevalent image format for tabular datasets. In contrast, the JSON file is excellent for storing "tree-like data."

Types of information in datasets

o Numerical data, such as a house's price, the temperature, and so on-Yes/No, True/False, Blue/Green, and other categorical data types are examples.

o Ordinal data: Ordinal data, like categorical data, can be quantified through comparison-A number of features in the real world can be difficult to maintain and process at first. As a consequence, pattern recognition methods have been suggested and use any dummy dataset.

o Dataset is Requested o To work on machine learning projects, we need a lot of data because ML/AI models can't be trained without it. The gathering and preparing of the dataset is one of the most critical components of developing an ML/AI project. The technology used in machine learning projects will not work well if the dataset is not properly prepared and pre-processed. During the construction of the ML project, the authors completely rely on the datasets. When constructing Learning applications, datasets are divided into twosections such as Training dataset and Test Dataset



To download the datasets, you'll need a fast internetconnection on your computer. The training set is a subset of the dataset that was used to train the machine learning model, and the results are already known.

The test set is a subset of the dataset used to test the machine learning model, and it is used by the model to forecast the output.

III. MACHINE LEARNING DATA PREPROCESSING

Preparing raw data for use in a classification model is known as data preparation. It's the most crucial and first step in creating a machine learning model. Whenever working on a machine learning project, we often do not have access to clean and prepared data. Also, before completing any data-related activity, the data must be cleaned and prepared. Asa response, we employ a data preprocessing procedure. The steps are as follows:

1. Obtaining the data
2. Adding Libraries
3. Datasets to be imported
4. Identifying Data Gaps
5. Categorical Data Coding
6. Creating a training and test set from the dataset
7. Scaling of features

3.1 Solutions for missing data

There are generally two options when dealing with missing data:

3.2 Eliminating a certain row: The first method istypically used when dealing with null values. We just delete the null values from the relevant row or column in this method. This technique, however, is inefficient, and deleting data may result in information loss, leading to incorrect results.

3.3 Obtaining the Mean: To fill in the gaps, we'll utilize the mean of the column or row, together with any blank values. This method works well for variables with a lot of quantitative data, such as age, salary, and year. This method will be used in this case.

3.4 Supervised Learning Steps:

o Determine thetype of training dataset first.

o Collect/gather the tagged training data

o Separate the training dataset into three sections: training, testing, and validation.

o Determine the input features of the training dataset, which should contain sufficient data for the model toaccurately predict the output.

o Select a suitable model algorithm, such as a support vector machine or a decision tree.

o Use the practice data to run the algorithm. Control parameters are occasionally necessary for validation sets,which are a subset of training datasets.

o Provide the test set for determining the model's accuracy. Our model is accurate if it correctly predicts the outcome.

supervised machine learning algorithms include thefollowing: 1. Regression - If there is a link betweentheinput and output variables, regression procedures are performed. It's used, among other things, to forecast continuous variables like weather and business patterns. Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, and Polynomial Regression are some well-known supervised learning regression methodologies. When the output variable is categorical, which implies it has two classes, such as Yes-No, Male-Female, True- False, and so on, classification methods are utilized.

There are two sorts of problems that theunsupervised learning algorithm can handle:

3.4.1. Clustering: Clustering is a method of groupingitems so that those with the most similarities areplacedin one group, while those with few or nosimilarities are placed in another. Cluster analysislooks for commonalities between data objects andcategorizes them depending on whether or not theyhave them.

3.4.2. Association: The association rule is an unsupervised learning method for detecting linksbetween variables in a large database. It recognizes the collection's set of objects that appear together. The association rule improvesthe efficiency of marketing campaigns. People who buy X (for example, a loaf of bread) are more likely to buy Y (butter/jam). Market Basket Analysis is a good example of an association rule.

3.5 Capabilities in Artificial Intelligence (AI) andMachine Learning

In practically every area, the convergence of AI and machine learning is revealing new opportunities. These are just a handful of the abilities that have been shown to help firms alter their processes and products.

3.5.1 Future-looking analytics

Speech recognition allows a computer system to recognize words in spoken language, whereas natural language understanding recognizes meaningin written or spoken language. These abilities enable the recognition of faces, objects, and actions in photos and videos, as well asthe implementation of functions like visual search. Retailers use AI and machine learning to optimize inventory, construct recommendation engines, and improve thecustomer experience with visual search, according to public opinion research. Artificial intelligence (AI) and machine learning (ML) are powerful cyber security technologies that help businesses discover anomalies and safeguard themselves and their customers. Customer happiness is extremely important**.**

3.6 Elimination by Reverse

Backward elimination is a feature selection method used in machine learning model construction. It's used to filter away features that don't have much of an effect on the dependent variable or outcome prediction. In Machine Learning, there are several methods for creating a model, including:

1. All-in
2. Reverse Elimination
3. Forward-thinking selection
4. Elimination in both directions
5. Comparison of Scores

We shall only utilize the Backward Elimination strategy in Machine Learning because it is the fastest.

3.6.1 Reverse Elimination Procedures

The following are the main steps in the backward elimination procedure:

To stay in the model, we must first select a significance threshold. (SL=0.05) Step 2: Fill in the model with all potentialpredictors and independent variables.

Step 3: Determine the P-value of the most important predictor: a. Proceed to step 4 if the P- value is greaterthan SL.

b. Otherwise, finish, and our model is complete.

Remove the predictor from the equation in step 4.

Step 5: Rebuild the model with the remaining variables and fit it.

3.6.2.Need for Backward Elimination: An optimal Multiple Linear Regression model:Going Backwards Is Required:

The Best Multiple Linear Regression Model:

In the previous chapter, we investigated and effectively created a Multiple Linear Regression model using four independent factors (R&D spend, Administration expense, Marketing spend, andstate (dummy variables)) and one dependent variable (R&D spend) (Profit). That model, however, isn't ideal because we've included all of the independent factors and have no idea which onehas the greatest influence on the forecast.

Unnecessary features add to the model's complexity. As a result, it is preferable to have onlythe most important elements and keep our model basic in order to achieve the best results. So, in order to improve the model's performance, we'llapply the Backward Elimination approach. This approach is used to improve the MLR model's performance by only including the most important features and excluding the least important ones. Let's apply it to our MLR modeling first.

IV.ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING RISKS

Astonishment with AI's speed and creativity is frequently followed by fear. Stephen Hawking, Elon Musk, and Bill Gates have all warned about the perils of artificial intelligence if it is not handled correctly by humans. Fears that computers will one day sprout their own brains have been fanned by popular fiction and movies. Some fear that harmful AI technologies, such as self-driving cars, will slip into the wrong hands. These worries aren't entirely unfounded. The usefulness of data mining algorithms in targeting social media users, as well as the consequences of tampering with technology, were demonstrated in the two most recent US presidential elections, for example. These interventions, on the other hand, were carried out by people who employed modern technology for nefarious purposes, rather than by sentient machines. Automation is a huge presence in our daily lives because of its convenience and pervasiveness, and it, like anything else, must be governed by legislation and ethics.

4.1 Cyber security is another subject to be concerned about. Cyber attacks are becoming more sophisticated and ingenious. Malware that uses artificial intelligence is the same as malware that uses any other artificial intelligence. AI is also working out how to cope with cyber security solutions that are based on AI. We're approaching the day where cyber security will be a battle between good and evil machines. Machine learning algorithms, on the other hand, are quite good at spotting anomalies. Cyber security specialists will have to constantly innovate to keep up with bad actors.

4.2 Data Science in the Future Artificial intelligence's learning mechanism is at its limit. Machines learn in stages to attain a certain result, basing future judgments on previous evidence. Humans, on the other hand, can reason abstractly, use context, and unlearn information that is no longer useful.

V. CONCLUSION

Future machine learning will benefit as a result of this. Digital assets such as financial and personal data are managed using algorithms. This could be the next step in enhancing AI security and reducingrisks. Even if robots are not entirely "intelligent" in the way humans think of intelligence, artificial intelligence will have a huge impact on the future of data science. Despite the fact that computers can process data quicker than humans, we have yet to develop software that can mimic our creative and logical abilities. Machines are useful, but they are merely an adjunct to human inventiveness. As we get closer to making science fiction a reality, deep learning and reinforcement learning are predictedto see breakthroughs in AI.

References

[1]. Adadi, A., &Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6,52138–52160. https://doi.org/10.1109/AC CESS.2018.2870052.

[2]. Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., &Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. International Journal of Hospitality Management, 80, 5277. https://doi.org/10.1016/j.ijhm.2019.01.003 .

[3] Amorós, L., Hafiz, S. M., Lee, K., &Tol, M. C. (2020). Gimme that model!: A trusted ML model trading protocol. arXiv:2003.00610 [cs]. http://arxiv.org/abs/2003.00610 4. Assaf, R., & Schumann, A. (2019).

[4]. Explainable deep neural networks for multivariate time series predictions. Proceedings of the 28th International Joint Conference on Artificial Intelligence, 6488–6490. https://doi.org/10.24963/ijcai.2019/932 .

[5]. Bastan, M., Ramisa, A., &Tek, M. (2020). Cross-modal fashion product search with transformer-based Embeddings. CVPR Workshop - 3rd workshop on Computer Vision for Fashion, Art and Design, Seattle: Washington.

[6]. Brynjolfsson, E., & McAfee, A. (2017).The business of artificial intelligence. Harvard Business Review, 1–20.

[7]. Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., & Song, D. (2018). Robust physical- world attacks on deep learning visual classification. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018,1625– 1634. https://doi.org/10.1109/CVPR .2018.00175 .

[8]. Halidu, M. K., Bagheri-Zadeh, P., Sheikh-Akbari, A., &Behringer, R. (2019, June).Pca in the context of face recognition with the image enlargement techniques.In 2019 8th Mediterranean Conference on Embedded Computing (MECO) (pp. 1–5).

[9]. Wijaya, I. G. P. S., Husodo, A. Y.,&Arimbawa, I. W. A. (2016, October).Real time face recognition using DCT coefficients based face descriptor. In 2016 International Conference on Informatics and Computing (ICIC) (pp. 142–147).

[10]. Jiang, C., Wang, M., Tang, X., & Mao, R.(2019, November). Face recognition method based on sparse representation and featurefusion. In 2019 Chinese Automation Congress (CAC) (pp.396–400).

[11]. Xiao, J., &Weiwei, G. (2019, October).Face recognition algorithm based on Prewitt and convolutional neural network. In 2019 3rdInternational Conference on Electronic Information Technology and Computer Engineering (EITCE) (pp.2007–2009).

[12]. Kerımbekov, Y., & Bılge, H. Ş. (2017,May). Face recognition via Lorentzian metric. In 2017 25th Signal Processing andCommunications Applications Conference (SIU) (pp. 1–4).

[13]. Chang, L., Yang, J., Li, S., Xu, H., Liu,K., & Huang, C. (2018, November). Face recognition based on stacked convolutionalautoencoder and sparse representation. In 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP) (pp. 1–4).

[14]. Nandhini, R., Duraimurugan, N.,&Chokkalingam, S. P. (2019). Face recognition based attendance system. International Journal ofEngineering and Advanced Technology (IJEAT), 8(3S). 15. Khan, M., Chakraborty, S., Astya, R.,&Khepra, S. (2019, October). Face Detection and Recognition Using OpenCV. In 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 116–119).

[15]. Khan, M., Chakraborty, S., Astya, R., &Khepra, S. (2019, October). Face Detection and Recognition Using OpenCV. In 2019International Conference on Computing,Communication, and Intelligent Systems (ICCCIS) (pp. 116–119).