



Enhancing Predictive Modeling with Human-Like Intelligence: A Deep Feature Synthesis Approach

Prosant Kumar Mahanty^{1*}, Dr. Anoop Sharma²

1. Research Scholar, University of Technology, Jaipur, Rajasthan, India
mahantyprosant@gmail.com ,

2. Professor, Department of Computer Science & Engineering, University of Technology, Jaipur, Rajasthan, India

Abstract: In this work, we create the Data Science Machine, an automated tool for extracting predictive models from unprocessed data. We initially propose and build the Deep Feature Synthesis method for automatically creating features for relational datasets in order to accomplish this automation. The pursuit of Human-Like Intelligence (HLI) in AI systems, the creation of emotionally intelligent AI, and the possible convergence of XAI with cognitive sciences are all further explored in this study. The advancement of artificial intelligence (AI) towards Artificial General Intelligence (AGI) raises important questions about consciousness, ethics, and social consequences.

Keywords: Deep Features synthesis, Predictive modeling, artificial Intelligence, HLI

----- X -----

INTRODUCTION

Although predictive analytics is an essential tool for corporate decision-making, we may get even more precise and useful insights by integrating AI models. At Key Labs, we are aware of the potential effects that AI-driven forecasts might have on companies in a range of sectors. With the help of machine learning and artificial intelligence approaches, we can create complex algorithms, analyse enormous volumes of data, and make predictions rapidly thanks to AI models. We can improve our skills in data analysis, algorithm creation, and predictive modelling by using AI in predictive analytics. Predictive analytics has been transformed by AI, which enables companies to swiftly process and evaluate enormous volumes of data. AI models, like machine learning algorithms, allow historical data to be mined for important insights and trends. Accurate forecasting of future events is made possible by these insights, providing businesses with a competitive edge and enhancing decision-making in a variety of business domains.

Optimising company processes is one of the main advantages of using AI into predictive analytics. AI models, for instance, may improve inventory management via the analysis of historical sales, consumer behaviour, and market trends data. Businesses may optimise inventory levels, save expenses, and lower the risk of stockouts or overstocking by precisely forecasting demand trends. Optimising marketing initiatives also benefits from the use of AI-powered predictive analytics. AI algorithms may provide customised suggestions and targeted ads by examining data on client demographics, preferences, and behaviour. This raises consumer satisfaction and conversion rates while simultaneously enhancing the efficacy of marketing initiatives.

AI models may also be used to streamline logistical processes and delivery routes. Businesses may identify the most effective routes and delivery schedules by analysing data on traffic patterns, weather, and consumer locations. These results in lower environmental impact, increased consumer happiness, and cost savings. Predictive analytics benefits greatly from AI's capacity to manage large, complicated datasets and derive insightful information. Businesses may improve client experiences, increase operational efficiency, and make data-driven choices by using AI models.

LITERATURE REVIEW

Abdulaziz Aldoseri et.al (2023) Artificial intelligence (AI) is finding increased use in a variety of sectors, including banking, healthcare, and transportation. Large-scale dataset analysis is the foundation of artificial intelligence, necessitating a steady stream of high-quality data. Nonetheless, there are difficulties with using data for AI. The difficulties of utilising data for AI are thoroughly reviewed and critically examined in this study, along with issues related to data amount, quality, privacy, security, bias, and fairness, as well as technological know-how and abilities. This essay looks closely at these issues and makes suggestions for how businesses and organisations can deal with them. Organisations may use AI to make better selections and get a competitive edge in the digital era by comprehending and tackling these issues. It is anticipated that the scientific research community will find this review article very beneficial in developing fresh and original ideas to reconsider our methods for approaching data strategies for AI, given that it presents and analyses a variety of strategies for data difficulties for AI throughout the last ten years.

Christopher Collins et.al (2021) In recent years, the information systems (IS) research community has given artificial intelligence (AI) more attention. But there's a growing worry that, like with IS research in the past, there may not be enough cumulative buildup of knowledge in AI research. By completing a thorough literature assessment of AI research in IS from 2005 to 2020, our work tackles this question. 1877 papers were found using the search method; 98 of these were found to be primary studies, and a summary of the major topics that are relevant to this research is provided. This study offers significant insights into the present stated commercial value and contributions of artificial intelligence (AI), as well as practical consequences for AI usage and research. Additionally, the study identifies potential for future AI research in the form of a research agenda.

Debleena Paul, et. al (2021) The pharmaceutical business is among the first industries to benefit from artificial intelligence (AI), which has only begun to expand its applicability across a number of other sectors. The impactful application of AI in various pharmaceutical domains, such as drug development and discovery, drug repurposing, increasing pharmaceutical productivity, clinical trials, etc., is highlighted in this review. This reduces the workload for humans and expedites the achievement of goals. Future prospects for AI in the pharmaceutical sector are also covered, as well as crosstalk on the methods and instruments used to enforce AI and the current difficulties that need to be solved.

James Max Kanter et.al (2015) In this work, we create the Data Science Machine, an automated tool for extracting predictive models from unprocessed data. We initially propose and build the Deep Feature Synthesis method for automatically creating features for relational datasets in order to accomplish this automation. The method creates the final feature by successively applying mathematical functions along a route that follows relationships in the input to a base field. Second, we use a new process-based technique

based on Gaussian Copula to design and modify a generalizable machine learning pipeline. In three data science challenges with 906 other data science teams, we joined the Data Science Machine. We outperform 615 teams in these data science contests with our method. We outperformed most of the competitors in two out of the three events, and in the third, we scored 94% of the highest-ranking competitor. In the best-case scenario, we defeated 85.6% of the teams and scored 95.7% of the highest among the submitted entries in the tournament.

José Jiménez-Luna, et.al (2020) Deep learning has potential applications in drug development, such as sophisticated image analysis, molecular structure and function prediction, and automated synthesis of novel chemical entities with customised features. The underlying mathematical models are often difficult for the human mind to understand, despite the increasing number of promising applications. To meet the requirement for a new narrative of the machine language of the molecular sciences, there is a need for "explainable" deep learning techniques. The most important algorithmic ideas in explainable AI are outlined in this review, along with future prospects, possible uses, and a number of unmet difficulties. Additionally, we hope that it inspires further work towards the creation and use of explainable AI methods.

DEEP FEATURE SYNTHESIS: ARTIFICIAL INTELLIGENCE

The input properties of a machine learning system have a major impact on its effectiveness. The objective of feature engineering is to turn unprocessed data into meaningful characteristics that a machine learning system can exploit. When coming up with features, data scientists usually draw on their prior experience or specialised expertise. However, the number of ideas that may be examined and the order in which they can be studied are limited by the time and effort required to develop a feature concept. Because of this, data scientists must decide which ideas to test first and when to give up on them so they may concentrate on other aspects of their jobs. Decomposing the characteristics that data scientists create into the generalised procedures they adhere to produces Deep Feature Synthesis. This chapter presents the Deep Feature Synthesis pseudocode, explains the feature generating abstractions, and explains the reasoning behind the technique.

Deep Feature Synthesis algorithm

A collection of related things serves as the input for Deep Feature Synthesis. Every entity that the table is built on has a primary key, which serves as a unique identifier for each instance of the entity. An entity may have an optional foreign key that uniquely identifies a related entity instance. Fields in an instance of the entity may be classified as numeric, category, timestamp, or free text data types.

Notationally, we have entities supplied for a certain database by $E1\dots$, These fields are included in each entity table and are indicated by $x1\dots|Ej|$. \vec{x}_{ij} shows the range of values for field i in the given object. An instance of the entity is represented by each value in this array j . \vec{f}_{ij} symbolises an array containing feature values created from field i values in entity j .

LEARNING FROM THE BRAIN TO ENHANCE AI

Neuroscientific principles in XAI The area of XAI is expanding along with the need for openness in AI systems (Adadi & Berrada, 2018). The decision-making processes of complicated models have been made

more understandable by methods like Shapley Additive Explanations (SHAP) (Lundberg & Lee, 2017) and Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016). Furthermore, techniques like Integrated Gradients (IG) (Sundararajan et al., 2017) and Layer-wise Relevance Propagation (LRP) (Bach et al., 2015) provide more in-depth understanding of the impact of input characteristics in certain model designs. These techniques bear resemblance to the ways in which neuroscientists try to interpret neural network activity inside the brain.

Incorporating neuroscientific insights Recent multidisciplinary research have emphasised the difficulty of integrating neuroscience results into AI development because of the complexity of brain systems (Marcus, 2018; Hassabis et al., 2017).

Bridging research fields Recent joint initiatives have shown that there is still a considerable barrier in bridging techniques and terminologies between AI and neuroscience (Yamins & DiCarlo, 2016; Kriegeskorte, 2015).

Neurodiversity and XAI adaptability The adaptability of XAI is evaluated in a number of fields, each with specific interpretability criteria. Explainable models, for example, may help medical professionals comprehend diagnoses made with AI assistance (Holzinger et al., 2017). In the financial domain, they may assist in providing suppliers and customers with clarification on credit rating algorithms (Weerts et al., 2023). Furthermore, XAI is playing a bigger part in domains where choices have a big impact on sustainability and safety, such environmental modelling and autonomous driving. It is critical to adapt XAI technologies to these various requirements in order to ensure their successful integration across various industries.

Real-world XAI applicability Real-world settings are where XAI approaches need to demonstrate their value (Baniecki et al., 2021). Particularly when working with big and sophisticated AI systems, their applicability and scalability are essential (Hedstrom et al. ", 2023). Widespread use of scalable XAI systems is necessary to guarantee that interpretability does not result in decreased performance or higher resource requirements.

AI visualization and brain imaging Saliency Maps and Gradient Input are two visualisation techniques that are essential to improving the interpretability of AI models (Simonyan et al., 2013; Shrikumar et al., 2017). These techniques are also used in brain imaging (e.g., fMRI, PET scans) to determine which brain regions are active during particular tasks, providing a window into the brain's decision-making processes.

Cognitive neuroscience in XAI evaluation Sensitivity analysis is often used to evaluate the robustness of XAI approaches (Baehrens et al., 2010). Additionally, user studies are becoming a crucial component of assessing XAI efficacy, particularly concerning the ways in which various stakeholders understand and engage with AI explanations (Abdul et al., 2020). This examination aids in figuring out

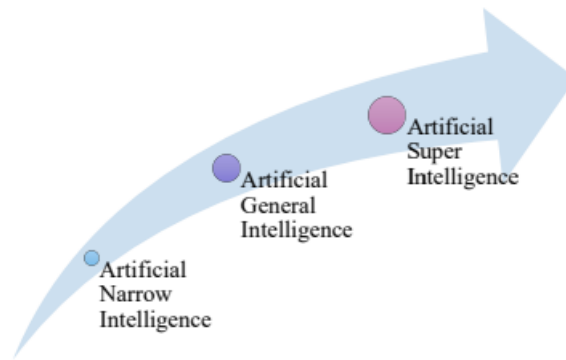


Figure 1: The evolution of artificial intelligence. The arrow's ascent reflects the dynamic growth and expanding capabilities of AI, marking key developments from basic algorithms to advanced sentient systems

PREDICTIVE MACHINE LEARNING PATH

We develop a generalised machine learning route to use the features generated by Deep Feature Synthesis.

Creating a prediction issue is the first stage. This is accomplished by choosing a dataset feature to model. This characteristic that we want to forecast is known as the target value.

We put together the characteristics that are suitable for use in prediction once a target value has been chosen. These attributes are known as predictors. Predictors are ruled out as invalid if they depend on data that was unavailable when the target value occurred, or if they are calculated using common base data as the target value. A database containing the information connected to every entity-feature is likewise kept up to date by the Data Science Machine. This metadata includes details on the original database's base fields that were used to create the feature, along with any temporal dependencies that could have been present.

A. Reusable machine learning pathways

The Data Science Machine offers a parametrized approach for data preparation, feature selection, dimensionality reduction, modelling, and assessment when a target feature and predictors have been chosen. To adjust the settings, the Data

Table 1: Summary of Parameters in the Machine Learning Pipeline and the Optimized Values from Running GCP by Dataset

Param	Default	Range	Function	KDD cup 2014	IJCAI	KDD cup 2015
k	1	[1, 6]	The number of clusters to make	1	1	2
n_c	100	[10, 500]	The number of SVD dimensions	389	271	420
γ	100	[10, 100]	The percentage of top features selected	32	65	18
rr	1	[1, 10]	The ratio to re-weight underrepresented classes	4	8	1
n	200	[50, 500]	The number of decision trees in a random forest	387	453	480
m_d	None	[1, 20]	The maximum depth of the decision trees	19	10	10
β	50	[1, 100]	The maximum percentage of features used in decision trees	32	51	34

Science Machine offers an intelligent parameter optimisation tool. The procedures for creating prediction models and machine learning are as follows:

Data preprocessing: We clean the data by eliminating the null values, transforming the categorical variables using one-hot encoding, and normalising the features before going on to the machine learning

route.

Feature selection and dimensionality reduction: Each object produced by Deep Feature Synthesis has several characteristics. We use two methods in order to decrease the size of the feature space: Using the Truncated SVD transformation, we first identify the n_c components of the SVD. Next, we compute the f -value of each SVD feature in relation to the target value, and we choose the $\eta\%$ top ranking features.

Modeling: We build n decision trees to create a random forest for modelling. Every decision tree employs a portion of the characteristics indicated by β and has a depth of md . It might be useful to have a distinct model for each cluster of data points in several datasets. We use a k means clustering approach to divide the training points into k clusters in order to include this. Then, for every cluster, we train a different random forest. A trained cluster classifier applies the associated model to a data point and assigns a cluster label in order to predict a label for a test sample.

There are situations where one of the target value classes in classification issues is underrepresented. The modelling step may re-weight an underrepresented class by a factor of rr to make up for this.

Four parameters have been incorporated during the modelling stage: n , md , β , k , and rr . We next go over how to automatically adjust this route.

Human-Like Intelligence

In AI research, Human-Like Intelligence (HLI) explores consciousness, self-awareness, and emotional intelligence with the goal of endowing robots with human-like cognitive and affective capacities (Assran et al., 2023). In order to facilitate informed decision-making, the study of consciousness entails building self-aware systems, which presents technological difficulties as well as philosophical and ethical issues (Dehaene, 2014; Tononi et al., 2016). Assran et al. (2023) explain how HLI integrates context-aware computing, emotion identification, and natural language processing to emphasise emotionally intelligent AI. This fosters natural, empathic connections and has transformational potential in areas like mental health, customer service, and personal support. The combination of emotional intelligence and awareness in AI represents a major advancement in the replication of human complexity. Beyond technical details, the development of HLI involves a multidisciplinary investigation of human awareness, emotional reactions, and mental processes. The goal of this confluence of AI, neuroscience, psychology, and ethics is to develop AI systems that completely understand and connect to the human experience, in addition to simulating it (Dehaene, 2014).

CONCLUSION

The Data Science Machine is a comprehensive system designed to perform data science operations on relational data. Deep Feature Synthesis, a technique for autonomously creating features for machine learning, lies at the heart of it. Through the use of autotuning, we are able to optimise the whole pipeline without the need for human intervention, allowing it to generalise across various datasets. Furthermore, the investigation of Human-Like Intelligence (HLI) in AI creates new avenues for emotional intelligence and human-like interactions, influencing AI's future course towards interpretability, transparency, and a more nuanced comprehension of the human condition.

References

1. Aldoseri, A.; Al-Khalifa, K.N.; Hamouda, A.M. Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Appl. Sci.* 2023, 13, 7082. <https://doi.org/10.3390/app13127082>
2. Christopher Collins et.al “Artificial intelligence in information systems research: A systematic literature review and research agenda” *International Journal of Information Management* Volume 60, October 2021, 102383
3. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today.* 2021 Jan;26(1):80-93. doi: 10.1016/j.drudis.2020.10.010. Epub 2020 Oct 21. PMID: 33099022; PMCID: PMC7577280.
4. Kanter, J.M., & Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1-10.
5. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell* 2, 573–584 (2020). <https://doi.org/10.1038/s42256-020-00236-4>
6. J. Wang, Z. Wang Strengths, Weaknesses, Opportunities and Threats (SWOT) Analysis of China’s Prevention and Control Strategy for the COVID-19 Epidemic *Int J Environ Res Public Health*, 17 (7) (2020), p. 2235, 10.3390/ijerph17072235
7. Koerhsen, W. (2018, June 2). Automated Feature Engineering in Python. Retrieved from <https://towardsdatascience.com/automated-featureengineering-in-python-99baf11cc219>
8. Kanter, J. M. and Veeramachaneni, K., "Deep feature synthesis: Towards automating data science endeavors," 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015, pp. 1-10, doi: 10.1109/DSAA.2015.7344858.
9. Patel, K. (2020, September 9). Image Feature Extraction: Traditional and Deep Learning Techniques. Retrieved from <https://towardsdatascience.com/image-featureextraction-traditional-and-deep-learning-techniquesccc059195d04>
10. Zahnert Thomas. (2011). The differential diagnosis of hearing loss. *Deutsches Arzteblatt international.*, 108(25): 433-43
11. Waleed B. Alshuaib, Jasem M. Al-Kandari and Sonia M. Hasan (2015). Update On Hearing Loss, Intech Open Book, Chapter 2, ISBN 9789535121954
12. Zeineb Abdmouleh, Adel Gastlia, LazharBen Brahim, Mohamed Haouari, Nasser AhmedAl Emadia, 2017, Review of optimization techniques applied for the integration of distributed generation from

renewable energy sources, *Renewable Energy*, 113: 266-280.

13. Mahendra Tiwari, Randhir Singh and Neeraj Vimal. (2013). An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education. *International Journal of Computer Science and Mobile Computing*, 2(2):53 – 57.
14. Sherine Dominick and Abdul Razak, T. (2014). Analyzing the Student Performance using Classification Techniques to find the better Suited Classifier. *International Journal of Computer Applications*, 104(4):1-3.
15. Abdullah AL-Malaise, Areej Malibari and Alkhozae. (2014). Students' Performance Prediction System Using Multi Agent Data Mining Technique. *International Journal of Data Mining and Knowledge Management Process.*, 4(5):1-20.