SIG788

Engineering AI solutions

Pass Task 2

Arunkumar Balaraman
S223919051

Contents

Target = Pass

Part 1: Paper Software Engineering for Machine Learning

Introduction

The integration of machine learning (ML) into software engineering processes presents new challenges for teams. Unlike traditional software components, ML models require specialized skills for data management, model customization, and handling as modules. Data discovery and management complexities, the need for diverse skill sets, and challenges in treating ML components as distinct modules highlight the unique considerations in building ML applications. As organizations strive to leverage AI capabilities, understanding and addressing these fundamental differences are crucial for successful integration and deployment of ML technologies in software development workflows.

1. Review the attached paper and describe the main fundamental differences to building applications and platforms for training and building applications based on ML than we have seen prior in application domains. Summarize it into 3 points

   Data discovery and management: In machine learning, the quality and quantity of data play a crucial role in the performance of models. Data preprocessing, cleaning, and feature engineering are essential steps that significantly impact the effectiveness of ML algorithms. Moreover, the diversity and heterogeneity of data sources in ML projects require teams to implement robust data management practices to ensure data integrity and consistency throughout the model lifecycle.

   Skillset Requirements: In addition to software engineering and machine learning expertise, ML designers need proficiency in statistical analysis, data visualization, and domain knowledge to develop effective models. Understanding the underlying algorithms, hyperparameter tuning, and model evaluation techniques are critical skills for optimizing model performance. Collaborative teamwork involving data scientists, domain experts, and software engineers is often necessary to address the multidisciplinary nature of ML projects.

   Module Boundaries: The interconnected nature of ML models poses challenges in maintaining modular boundaries within software systems. As models interact with each other and share dependencies, changes in one model can have cascading effects on others, leading to complex entanglements. Ensuring proper encapsulation and abstraction of ML components becomes essential to manage dependencies and facilitate scalability and maintainability in ML applications. Adopting modular design principles and clear interfaces can help mitigate the complexities arising from intertwined ML modules.

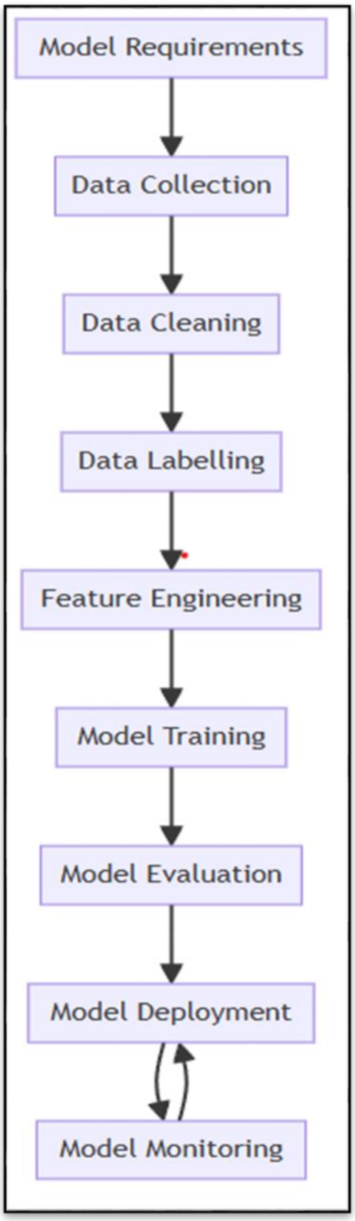
Citation:

Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (1, 5, 8, 9)

2. What are the main stages of machine learning workflow and explain each stage briefly. (Minimum 500 words)

Model Requirements: This is the first step where the problem is defined and the goals of the ML model are set. It involves a collaborative effort between stakeholders and data scientists to articulate business needs, identify data sources, and establish the success metrics for the model.

FlowChart :



Data Collection: This stage involves gathering the necessary data to train the model. It could involve collecting new data, identifying existing data within the organization, or sourcing data from external databases. The aim is to gather a dataset that accurately represents the problem domain and includes all relevant variables for the model to learn from.

Data Cleaning: Also known as preprocessing, this stage involves removing or correcting inaccuracies, inconsistencies, and outliers in the data. It also includes handling missing values and duplicate entries. Clean data is vital for training reliable and accurate models, as the quality of the data directly influences the model's performance.

Data Labeling: In supervised learning, each instance in the dataset must be labeled with the correct outcome. This stage involves tagging data with these outcome labels, which the model will learn to predict. This can be a labor-intensive process, particularly for large datasets, and may require domain experts to ensure accuracy.

Feature Engineering: This process involves selecting, modifying, or creating new features from the raw data to enhance model performance. It requires domain knowledge to identify what information is relevant to the problem and how it can be structured in a way that the ML model can understand and learn from.

Model Training: This is the stage where the chosen algorithm learns from the data. The algorithm is fed the training data, and it learns to map the input features to the desired output. The training process involves adjusting the model parameters to minimize the difference between the predicted and actual outcomes.

Classification Models: Logistic Regression, Random Forest, etc..
Regression Models: Linear Regression, Random Forest, etc..
Clustering: Density-based, Centroid-based, Hierarchical-based, etc..

Model Evaluation: After training, the model's performance is evaluated. This involves using a separate dataset (validation or test set) that the model hasn't seen during training. Evaluation metrics such as accuracy, precision, recall, and F1 score are used to assess how well the model meets the initial requirements and objectives.

Model Deployment: Once the model meets the performance criteria, it is deployed into a production environment where it can start making predictions or decisions based on new data. Deployment strategies can vary, including real-time inference, batch processing, or embedding the model into an application.

Model Monitoring: After deployment, continuous monitoring is crucial to ensure the model performs as expected over time. Monitoring involves tracking the model's predictive performance, detecting any significant changes in the data it's processing, and identifying opportunities for further improvement or retraining.

Each stage in the ML workflow plays a pivotal role in the development of effective machine learning models. By systematically following these steps, can build, deploy, and maintain models that provide valuable insights and inform decision-making processes.

3. Which domains within the Microsoft team have employed AI, and what machine learning approaches have they utilized?

Microsoft team have employed AI in different areas. The document mentions that AI is used in traditional areas such as

- search,
- advertising,
- machine translation,
- predicting customer purchases,
- voice recognition, and
- image recognition

Additionally, AI is being utilized in novel areas such as identifying

- customer leads,
- providing design advice for presentations and
- word processing documents,
- offering unique drawing features,
- healthcare applications, and
- improving gameplay

In terms of machine learning approaches utilized by these domains, states that a broad spectrum of machine learning approaches is being used, including **classification**, **clustering**, **dynamic programming**, **statistics**, **user behavior modeling**, **social networking analysis**, and **collaborative filtering**.

Different divisions within the company specialize in various aspects of machine learning, such as ranking and relevance algorithms, query understanding, natural language processing, sentiment analysis, intent prediction, summarization, machine translation, ontology construction, text similarity, and connecting answers to questions.

Other areas like Finance and Sales focus on building risk prediction models and forecasting, while internal resourcing organizations make use of decision optimization algorithms for resource optimization, planning, pricing, bidding, and process optimization **(Page 4)**.

This demonstrates the diverse applications of AI and the wide range of machine learning approaches being employed across different domains within the Microsoft team.

Citation:

Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (4,5)

4. What are the best practices with ML in software engineering? Provide a summary for each practice. (minimum 600 words)

Key challenges in developing large-scale ML applications and platforms, and how they are tackled in various products. These challenges were identified through card sorting of interview and survey responses. As software engineering and AI researchers, we then highlighted the most critical challenges for AI integration in software teams.

End-to-end pipeline support:

End-to-end pipeline support is crucial for integrating machine learning (ML) development into traditional software engineering processes. It involves creating a seamless development experience that covers all stages of ML development, including data collection, preprocessing, model training, evaluation, deployment, and monitoring. By automating the training and deployment pipeline and integrating model building with the rest of the software, teams can ensure smooth system deployment. Common versioning repositories for both ML and non-ML codebases and tight coupling of ML and non-ML development sprints and standups are recommended to facilitate fast-paced model iterations.

Data availability, collection, cleaning, and management:

Data availability, collection, cleaning, and management are essential for successful ML projects. Teams need to ensure that relevant datasets are collected from various sources, both internal and external. Data cleaning involves preparing the dataset by handling missing values, outliers, and inconsistencies. Proper data management practices, such as data versioning and sharing techniques, are necessary to manage the continuous changes in data sources and ensure data quality for training ML models.

Education and Training:

With the increasing ubiquity of machine learning in customer-facing products, engineers with traditional software engineering backgrounds need to learn how to work alongside ML specialists. Providing education and training opportunities, such as internal conferences on machine learning and data science, helps scaffold engineers' learning. Customized machine learning environments and tools, like Azure ML for Visual Studio Code and Azure ML Studio, enable engineers to build and deploy models effectively, catering to individuals with varying levels of experience.

Model Debugging and Interpretability:

Model debugging and interpretability are critical for understanding and improving the performance of ML models. Teams should focus on techniques that enhance model interpretability, such as feature importance analysis, SHAP values, and model visualization. By debugging models and interpreting their decisions, teams can identify and rectify issues like bias, overfitting, or underfitting, ensuring the model's reliability and effectiveness in real-world applications.

Model Evolution, Evaluation, and Deployment:

Model evolution, evaluation, and deployment are key stages in the ML lifecycle. Continuous evaluation of models using metrics like accuracy, precision, and recall helps teams assess model performance and identify areas for improvement. Model deployment involves integrating the trained model into production systems for making predictions on new data. Regular monitoring and updating of deployed models ensure their continued effectiveness and relevance in dynamic environments.

Compliance:

Compliance with principles of fairness, accountability, transparency, and ethics is essential when deploying AI systems in the open world. Teams should align their engineering practices and behaviours with these principles to ensure ethical AI and ML practices. Compliance with regulatory requirements and ethical guidelines is crucial for building trust with users and stakeholders and mitigating potential risks associated with AI applications.

Varied Perceptions:

The integration of machine learning components into applications is influenced by teams' prior experience with machine learning and data science. Teams with experienced data scientists and researchers may have different challenges and approaches compared to teams that are rapidly growing their expertise. Understanding and addressing the varied perceptions and challenges faced by different teams is essential for promoting knowledge sharing, collaboration, and continuous improvement in ML practices across the organization.

Citation:

Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (5,6)

Part 2: Fuzzy Logic: Application in Healthcare Supply Chain Management

Introduction:

The integration of fuzzy logic into numerous fields has transformed the decision-making process by effectively managing uncertainties and imprecisions. The healthcare sector, in particular, requires meticulous supply chain management due to the critical nature and immediacy of medical supplies and equipment. This report delves into the utilization of fuzzy technology to automate operations within the healthcare supply chain, thereby boosting efficiency, minimizing costs, and elevating the quality of patient care.

1. Target Domain and Application

The healthcare supply chain represents a crucial area where the efficient management of inventory directly influences patient care quality and operational costs. Fuzzy logic is applied in this domain to automate the ordering process, optimize stock levels, and forecast demand for medical supplies and pharmaceuticals accurately.

2. Fuzzy logic In Healthcare Supply Chain & Its advantages

Fuzzy logic seamlessly integrates into healthcare supply chain management by addressing the uncertainties and imprecisions inherent in managing supply and demand. Unlike binary logic systems that rely on precise inputs, fuzzy logic thrives in fluctuating environments by employing ranges of values to make nuanced decisions.

Advantages:

Handling Uncertainty: Fuzzy logic adeptly processes ambiguous or incomplete data, crucial for predicting medical supply needs amidst unforeseen demand spikes or seasonal variations.

Flexibility: The systems ability to incorporate expert insights into its decision-making process enables a more sophisticated approach to inventory management, reflecting the real-world complexities of the healthcare supply chain.

Improved Decision Making: Fuzzy logic enhances decision-making accuracy by systematically managing uncertainties, resulting in optimal inventory levels and minimized waste.

3. Pipeline and System Flow Using Fuzzy Technology

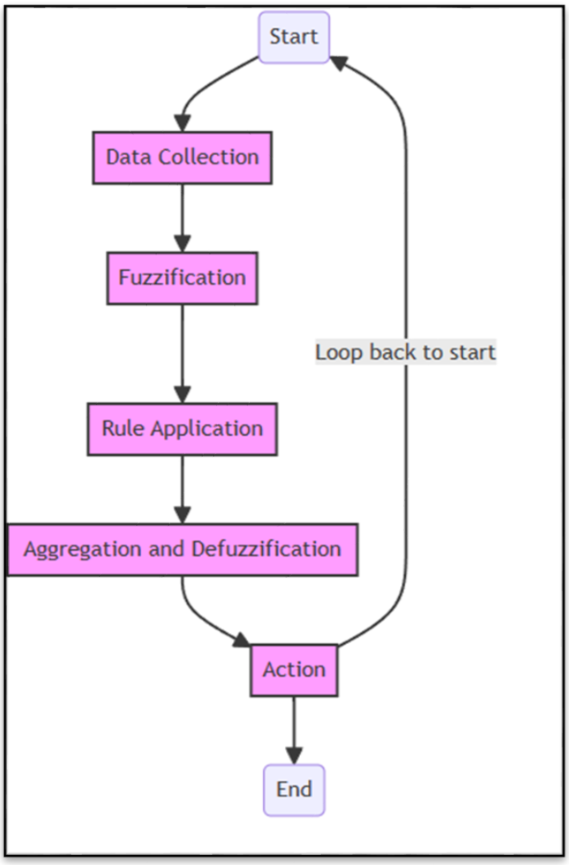
Fuzzification: The initial step is fuzzification, where quantitative variables such as inventory levels, demand predictions, and supplier reliability indicators are transformed into qualitative terms using membership functions. For example, numerical inventory levels are classified into fuzzy sets like "low," "medium," and "high." This transformation allows fuzzy logic to process precise data, accounting for the inherent uncertainty and variability in supply chain data.

Rule Application: During this stage, predefined fuzzy rules are applied to the fuzzified data. These rules are if-then statements that encapsulate expert knowledge and insights into the operation of the supply chain. For instance, a rule might state, "If the stock is low and the demand is high, then the reorder quantity is very high." By assessing the current state of the supply chain through these fuzzy rules, the system can make informed decisions amidst uncertainty.

Aggregation and Defuzzification: After the fuzzy rules are applied, the system aggregates the results from all relevant rules for a given decision point. This aggregated output is still in fuzzy terms and needs to be translated into a specific action. Defuzzification transforms this fuzzy output into a precise value or action, such as the exact number of items to reorder or the timing for the next order, facilitating practical decision-making.

Action: Based on the de-fuzzified outputs, the system carries out actionable tasks. This could include automatically placing orders with suppliers, adjusting inventory levels to align with anticipated demand, or updating delivery schedules. Automating these actions ensures that the healthcare supply chain operates efficiently, maintaining optimal stock levels to meet patient needs without incurring unnecessary costs or experiencing stockouts.

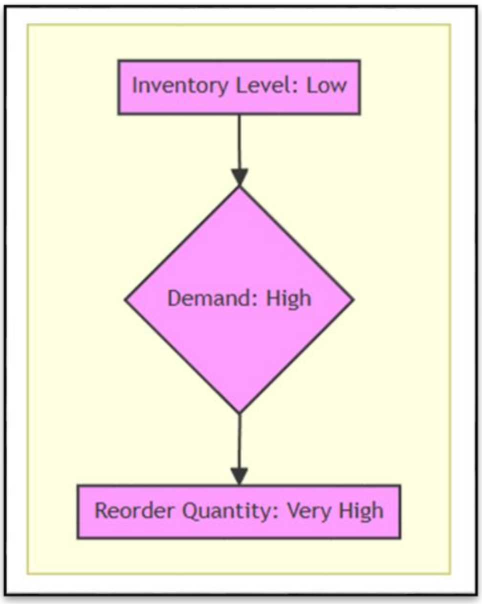
Pipeline Flow



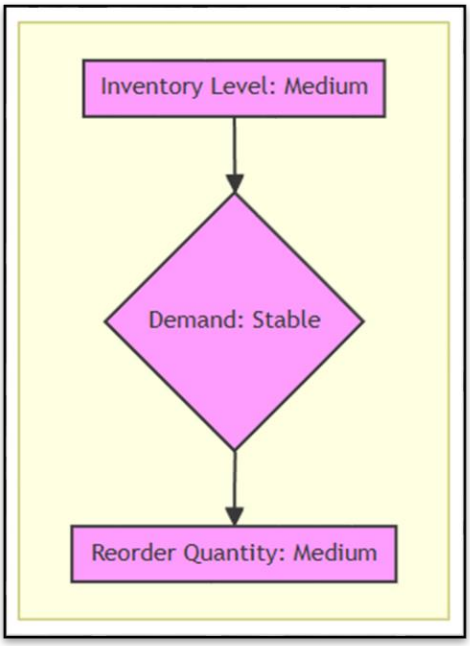
4. Fuzzy System Rules in the Healthcare Supply Chain

The fuzzy logic system within the healthcare supply chain might incorporate the following rules:

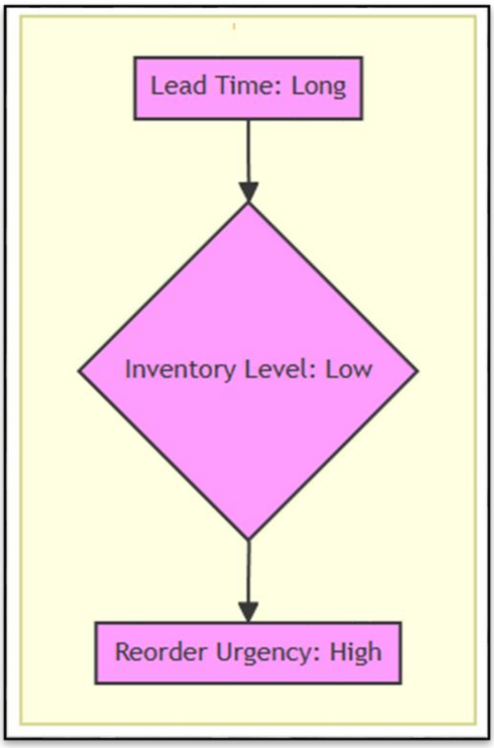
a. When the inventory level is “low” and the demand is “high,” the system determines that the reorder quantity should be “very high.”



b. If the inventory level is "medium" and the demand is "stable," the system suggests that the reorder quantity should be "medium."



c. If the lead time is "long" and the stock level is "low," then the reorder urgency is "high."



These rules employ linguistic variables to represent the status of the inventory and demand forecasts. This approach enables the system to make informed decisions even in situations where the data might be imprecise or incomplete. The use of fuzzy logic allows the system to interpret and operate on this data in a way that closely mimics human reasoning, making it particularly effective in managing complex systems like healthcare supply chains.

Conclusion:

Fuzzy logic's integration into healthcare supply chain management significantly addresses inherent challenges by managing uncertainties and imprecisions in supply and demand. This sophisticated approach enhances operational efficiency, reduces costs, and improves patient care quality.

Fuzzy logic not only automates key processes like ordering, stock level optimization, and demand forecasting but also offers substantial benefits. It can process ambiguous or incomplete data, ensuring medical supplies meet patient needs amidst unpredictable demand. The system's flexibility to incorporate expert insights allows for a nuanced approach to inventory management, reflecting the complexities of the healthcare supply chain. Furthermore, decision-making accuracy is enhanced, leading to optimal inventory levels, minimized waste, and judicious resource use.

The system flow, from fuzzification to action, underscores fuzzy logic's seamless integration into supply chain operations. Fuzzy logic translates quantitative data into qualitative actions through fuzzification, rule application, aggregation, and defuzzification, ensuring efficient supply chain management.

The implementation of specific fuzzy rules strengthens this process, enabling informed decisions based on the current state of the supply chain, thus mimicking human reasoning in managing complex systems.

In conclusion, fuzzy logic's deployment within healthcare supply chain management represents a progressive approach to tackling the sector's challenges. Its ability to reason under uncertainty is invaluable for ensuring the timely availability of critical medical supplies, enhancing patient care and operational efficiency. As healthcare evolves, the role of fuzzy logic in supply chain management will likely grow, paving the way for more innovative and effective future solutions.

References

- Great Learning. (2024). **Week 2 - Data Modeling**. Available at: https://olympus.mygreatlearning.com/courses/109553?module_id=747539

- Great Learning. (2024). **Fuzzy Logic - Control System**. Available at: https://olympus.mygreatlearning.com/courses/109553/pages/2-dot-7-fuzzy-logic-control-system?module_item_id=5377251

- Muhammet Deveci (17 July 2023) "**Effective use of artificial intelligence in healthcare supply chain resilience using fuzzy decision-making model**" Available at: https://link.springer.com/article/10.1007/s00500-023-08906-2

- Elizabeth Wallace (19 June 2023) "**Using Fuzzy Logic and AI to Improve Supply Chain Efficiency**" Available at: https://www.rtinsights.com/using-fuzzy-logic-and-ai-to-improve-supply-chain-efficiency/

Citations:

- Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (1, 5, 8)

- Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (4,5)

- Saleema Amershi., Andrew Begel, Christian Bird., Robert DeLine., Harald Gall., Ece Kamar., Nachiappan Nagappan., Besmira Nushi. & Thomas Zimmermann., Software Engineering for Machine Learning: A Case Study. Microsoft Research. Page (5,6)