Task 2 P: Case study on AI Solution and Fuzzy Logic

Part 1:

Ques-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.

Ans-1:

The research is about 'Software Engineering for Machine Learning: A Case Study', which shows how Microsoft incorporates Machine Learning with their software engineering practices. This research paper provides the knowledge about traditional software development and Machine Learning applications. Summarizing differences between building applications based on Machine Learning (ML) and traditional software development approaches:

- Data Management Complexity:
  ML heavy relies on data (both structured and non-structured data). Finding, cleaning, storing, and versioning data pose challenges for ML projects where else traditional software development deals with well-defined code.
- Customization and Reuse:
  Software development effective on reusability of code (like functions and libraries) where customisation in ML models is very complex (need pre-experience in ML with maths with expertise). In ML reusability of trained model for similar purpose is easy but significant changes are required for retraining or rebuilding the model.
- Modularity:
  Traditional software promotes modularity with clear boundaries between components for independent development whereas maintaining such boundaries in ML systems is difficult.

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

Ans-2:

Machine learning workflow is divided into 9 stages:

1. Model Requirement
   Designers make the call on which features can be practically brought to life using machine learning and which ones could enhance an existing product or a fresh one. Crucially, they also determine the most suitable types of models for the specific problem at hand.
2. Data Collection
   Teams seek out and blend accessible datasets, be it internal resources or those freely available through open sources, or they gather their own. Many times, they initiate training with a foundational model using generic datasets like MNIST, CIFAR, ImageNet and more for tasks such as object detection. Following this, they leverage transfer learning, amalgamating it with more tailored data to refine and specialize their model further, as seen in applications like pedestrian detection, car detection, face detection, and more.

3. Data Cleansing
   This task involves removing inaccurate or noisy records from the dataset, which is a common activity in all forms of data science. For these different libraries can be used, like Scikit-learn, Dask and more.
4. Data Labelling
   Assigning correct labels to each data entry. A majority of supervised learning methods necessitate labelled data for model training. Conversely, other approaches (like reinforcement learning) employ demonstration data or rewards from the environment to refine their strategies. Labels can be furnished by engineers, domain experts, or even crowd workers on online crowd-sourcing platforms.
5. Feature Engineering
   This refers to all the activities done to extract and select informative features for machine learning models. For certain models like convolutional neural networks, this stage is not so explicitly separate and often mixed with the next stage, which is model training.
6. Model Training
   The selected models (utilizing the chosen features) are trained and fine-tuned on the clean, gathered data along with their corresponding labels.
7. Model Evaluation
   The engineers assess the output model on validated or secure datasets employing pre-defined metrics.
8. Model Deployment
   The code that makes conclusions from the model is then put onto the specific device for use.
9. Model Monitoring
   The code that deduces outcomes from the model is then consistently watched for any potential errors while it runs in the real world.

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

Ans-3:

There are 3 Domains in which AI is being used within the Microsoft team:

1. Traditional Areas
   In traditional fields like Search and Advertising, machine learning has been extensively used to improve search result relevance, target ads better, and enhance user experience.
   AI is being employed in Machine Translation for real-time text, voice, and video translation, enabling communication across different languages and improving content accessibility.
   Machine learning is also being utilized in Customer Purchase Prediction to forecast customer buying behaviour, which helps in enhancing sales strategies and personalized marketing efforts.
   Voice Recognition technology is applied in virtual assistants and other services to convert spoken language into text, allowing for easier voice commands and dictation.
   Image Recognition is employed in various applications, ranging from photo categorization to object detection within images, which enhances user interaction and aids in data analysis.
2. Novel Areas:
   - Customer Lead Identification: Utilizing AI to scrutinize potential customer data for spotting and ranking leads, streamlining sales endeavours.
   - Design Assistance: Integrating AI into productivity tools to offer design suggestions in presentations and documents, enhancing visual appeal and user interaction.
   - Healthcare: Employing machine learning for diagnostic support, patient tracking, and customized treatment, advancing healthcare solutions. AI-powered chatbots assist healthcare providers with patient engagement, appointment scheduling, symptom checking, and triaging.

- Gaming: Elevating gameplay and user experience with AI, encompassing dynamic difficulty tweaks and enhancements in NPC behaviour. AI technologies are used for game development, player behaviour analysis, adaptive gameplay, and content recommendation on platforms like Xbox Live.

3. Infrastructure Project
   a. Incident reporting management:
      - Managing incidents and reporting them effectively.
      - Keeping track of unexpected events and documenting them appropriately.
   b. Identifying likely causes for bugs:
      - Figuring out the probable reasons behind software glitches.
      - Pinpointing the potential causes of errors in programs or applications.
   c. Monitoring fraudulent fiscal activity:
      - Keeping an eye on suspicious financial transactions.
      - Observing monetary activities for any signs of fraudulence.
   d. Monitoring network streams for security breaches:
      - Watching over data flows in networks to detect breaches in security.
      - Keeping a close check on network traffic to identify any unauthorized access or breaches in security protocols.

The following are the Machine Learning approach they have used in their organizations:

1. Classification and Clustering: Basic techniques for grouping data and making predictions based on input features, used across different areas for various purposes.
2. Dynamic Programming: Used to optimize problems and decision-making processes, like in gaming or operations research.
3. User Behaviour Modelling: Examining user interactions to foresee future actions, enhance user experience, and customize content.
4. Social Network Analysis: Making use of machine learning to comprehend and exploit the structures within social networks for suggestions, advertising, and more.
5. Collaborative Filtering: Employed in recommendation systems to anticipate user preferences by considering past interactions and resemblances with other users.
6. Natural Language Processing (NLP):
   - Finding out who or what is mentioned in text (Entity recognition)
   - Figuring out if the vibes in the text are positive, negative, or neutral (Sentiment analysis)
   - Predicting what someone's trying to do or say based on what they've written (Intent prediction)
   - Making long texts shorter while keeping the main points intact (Summarization)
   - Turning words from one language into words from another language (Machine translation)
   - Building maps of how words and ideas relate to each other (Ontology construction)
   - Seeing how alike two pieces of text are (Text similarity)
   - Making sure the answers match the questions (Connecting answers to question)

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

Ans-4:

Best practices have been mentioned below as per the report.

1. Education and training:

   As machine learning becomes more widespread in everyday products such as email and word processors, conventional software engineers must adapt to collaborate with machine learning specialists. At Microsoft, they facilitate the learning process for their engineers through various avenues. They conduct biannual conferences aimed at imparting the fundamentals of machine learning. Additionally, engineers deliver presentations on their tools and ongoing projects, while researchers share insights into the latest breakthroughs. Furthermore, teams hold weekly meetings to delve deeper into artificial intelligence topics. Moreover, there are online forums where individuals can pose queries and exchange knowledge on AI and machine learning.

2. Data availability, cleaning and management:

   Since machine learning depends heavily on large datasets, its success often relies on the availability, quality, and management of data. Labelling datasets is a time-consuming and expensive process, so ensuring they are accessible within the company is crucial. Reusing data helps minimize redundant efforts. Respondents prioritize data attributes like accessibility, accuracy, freshness, and connectivity, highlighting the importance of automation to efficiently gather data and create labelled examples, thereby accelerating experimentation with new models. Microsoft teams integrate data management tools with ML frameworks to avoid fragmentation in data and model management. Continuous changes in data sources require robust versioning and sharing techniques, such as tagging models with provenance information and datasets with extraction details. This facilitates the mapping of datasets to deployed models and improves data sharing and reusability.

3. Model debugging and interpretability:

   When we're addressing issues in components that learn from data, we don't merely search for coding errors. We also emphasize grasping the inaccuracies that occur due to the model's lack of precision or uncertainty. Determining the timing and causes of model errors is an actively researched domain. Certain individuals propose employing simpler models that are more comprehensible or devising methods to visualize complex models. When managing large systems that employ numerous models, individuals arrange them in a structured manner to facilitate the identification and resolution of errors.

4. Compliance:

   Microsoft has laid down some rules about how to use AI in real-life situations. These rules stress the significance of being fair, responsible, clear, and ethical. They want all Microsoft teams to make sure that their engineering practices and the way their software and services work follow these principles. Giving importance to these principles is very important in both software engineering and AI/ML processes.

5. Model evaluation and deployment:

   In the world of machine learning software, updates occur frequently due to changes in models, settings, and data. These alterations significantly impact the system's performance. Many teams employ both strict and flexible methods to test their experiments. They have structured approaches to experimenting with various combinations of changes, measuring multiple metrics to evaluate effectiveness, and occasionally manually verifying critical data. They maintain records of the changes made and their effectiveness. Automating tests holds paramount importance in machine learning, akin to traditional software. Teams meticulously design test suites to ensure their models function as intended. However, human oversight is also crucial in diagnosing issues when things go awry. Given the rapid pace of change, models require frequent updates. To facilitate smooth updates, engineers

advocate for automating the training and deployment process. They also stress the importance of centralized code management and close collaboration with other software teams.

6. End to end pipeline support:

   As machine learning progresses, blending it with software development becomes vital. Harmonizing this amalgamation poses challenges owing to variances between ML and conventional software. Nonetheless, it eases the workload and propels advancement. Users employ in-house systems or personalized configurations. A dependable data pipeline is imperative for smooth experimentation with AI algorithms. Automated pipelines simplify training, deployment, and integration, with lucid dashboards offering valuable insights. Platforms such as Azure ML for Visual Studio Code cater to engineers across proficiency levels, assisting in data management, model training, and upkeep. While visual aids aid novices, seasoned users might explore other options.

7. Varied Perceptions:

   We observed that how well product teams at Microsoft could blend machine learning components into their applications relied on their past exposure to machine learning and data science. Some teams had seasoned data scientists and researchers, while others had to quickly build up their expertise. Because of this mix in experience levels, we expected different challenges among the teams we surveyed.

   Two significant observations stood out. Firstly, irrespective of their experience, many respondents highlighted challenges linked to Data Availability, Collection, Cleaning, and Management as the most significant. Secondly, specific challenges varied in importance depending on the respondents' AI experience.

Part 2:

Application of Fuzzy logic: Loan Approval (India system) from BFSI domain.

What is the target domain and application?

The target domain for the application of fuzzy logic is the loan approval system. Mainly this system is governed by banks, NBFC and credit/co-operative societies, which relies over human touch. In traditional system, a human need to verify or cross-examine different tasks like KYC, savings account, credit score, purpose of loan, income tax return, income source and more, after which binary judgement is provided i.e., approve or reject. Fuzzy logic can automate the partial process by modifying human reasonings over borrower data by incorporating:

- Represent uncertainty
- Represent with degree
- Represent the belongings of a member of a crisp set to fuzzy set.

You need to explain how the fuzzy logic can fit into the target application. You need to discuss the advantage of using a fuzzy system in this domain/ application.

In traditional loan system, loan officer needs to use his experience and reasonings to verify different credit history, income and more, as its always not the case that data of borrower is full-fledged, there are some limitations or different problems also, like limited credit card history or income is not reliable. So, officer need to be strict with the thresholds to determine the loan to be provided or not.

Fuzzy logic allows us re-modelling the traditional system via using:

- Income can be divided into different slabs like, ‘low’, ‘medium’ or ‘high’.
- Defining fuzzy rules like, IF low income and high job security and credit card repayment on time then loan can be approved.
- Defining degree of membership for all factors like, income, credit score and more.

Advantage of using Fuzzy logic mechanism:

- Provides automatic way reasoning without much human interference, which can increase accuracy for different type of limitations in data.
- Fuzzy rules are easy to understand and showcase.

You need to provide the flow or pipeline of the system using fuzzy technology.

Step 1 - Input collection and define linguistic variables

Gathering relevant or pre-requisite information from loan applicant or borrower, like KYC, credit score, income, purpose of loan and more. Once information is collected and curated, then will define linguistic variables for each input variable, like ‘income’ -> ‘very-low’, ‘low’, ‘medium’, ‘high’, ‘very-high’.

Step 2 - Construct membership functions

Transforms the different inputs, which are crisp numbers, into fuzzy sets, assigning degrees of membership to each input variable.

Step 3 - Construct knowledge base rules and evaluate rules

Build a set of fuzzy rules into knowledge base in the form of IF-THEN-ELSE statements which will be used to evaluate the borrowers. Fuzzy rules which were constructed contains the relationship between input variables and output. Ex: IF low income and high job security and credit card repayment on time then loan can be approved.

Step 4 - Perform defuzzification

Defuzzification is then performed according to membership function for output variable. Here fuzzy output is converted back into a crisp value(percentage/score), that helps in loan decision.

Step 5 – Output decision

Based on defuzzied value, we can classify the loan output as approve, reject, or can't determine (need further checks).

You need to provide the rules that can be extracted from the fuzzy system in the domain / Application.

The fuzzy system in the loan approval system operates on different set of rules derived by finance experts, domain experts and analysing historical data. These rules are represented in the form of IF-THEN-ELSE statements. Some examples of rules created:

- IF (income is medium) AND (job security is high) AND (credit score is high) THEN (loan approval is high), with degree of truth as 0.9
- IF (income is low) AND (job security is low) THEN (loan approval is low), with degree of truth as 0.2
- IF (job security is medium) AND (income is medium) AND (credit score is medium) THEN (loan approval is medium), with degree of truth as 0.5

These are just few sample examples from large and complex rules which contains various factors for loan approval decision making.

Conclusion

Fuzzy logic offers a helping hand to loan officers, it won't eliminate the human touch but will be helpful in streamlining the process for loan approval system.