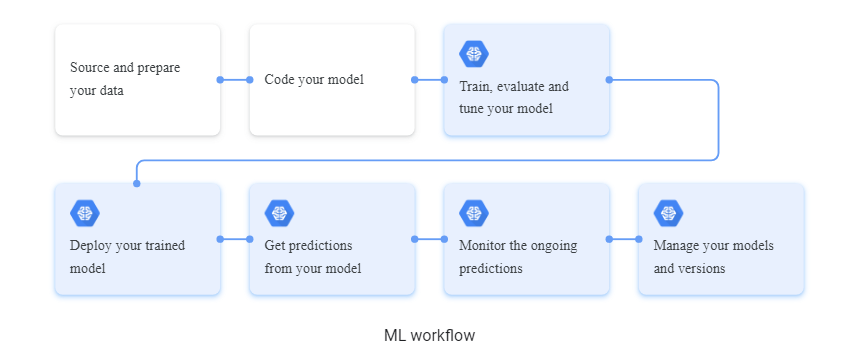
(i) What are the infrastructural foundations, best practices in the way of data science

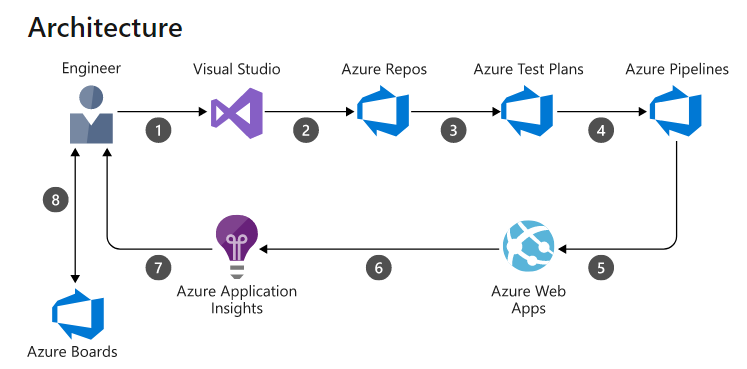
workflows that you would want to establish at first?

Ans –



Infrastructural Foundation

1. Serverless Cloud platform (Azure preferred).
2. Azure GPU NC6s for training.
3. Azure AKS for automatic load balance.
4. Azure DevOps for designing CI/CD pipeline.



(ii) What will be the hallmarks for you to consider switching from rules based to ML

models?

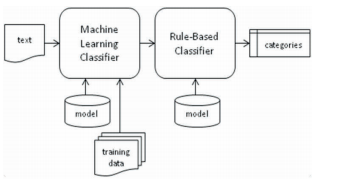
Ans –

1. ML enables real-time insights, identification of anomalies, and detection of subtle changes in patterns in large and dynamic data sets.
2. rules-based systems rely on pre-programmed rules to identify changes in behavior or predict outcomes. These rigid systems are not adaptable to an evolving and highly disrupted industry like financial services, which requires a more agile, flexible platform to overcome fraud challenges.
3. Rules-based approaches are time-intensive where an ML-based approach facilitates real-time processing.
4. rules-based requires manual work and supervision where ML enables automatic detection of anomalies; rules-based requires multiple steps for verification that impede the user experience where ML reduces and simplifies verification; and rules-based approaches identify obvious fraud patterns where ML is able to discover hidden fraud buried within subtle pattern changes.
5. most important difference between rules-based systems and ML is the cost factor. Rules based are far less scalable than ML models.

(iii) Would you consider hybrid approaches b/w rules based and ML model? If so,

which modules would be best suited for ML modeling?

Ans – Yes, Vector Space Model



1. Train a base model from scratch by using available training data (labeled corpus).
2. Vector dimensionality may be reduced using any of the Feature Selection techniques that have been proposed in the information retrieval field. These rules are used for post-processing the output of the machine learning classifier.
3. Fine-tuned output.

(iv) What are the user provided, 3rd party verified, publicly available, mobile & social

data sources that you would like to consider, and at which stages of the business?

Ans –

publicly available Data

<https://www.openml.org/d/1597>

<https://www.kaggle.com/mlg-ulb/creditcardfraud> from <https://mlg.ulb.ac.be/wordpress/>

<https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)> from <https://archive.ics.uci.edu/ml/index.php>.

Business Consideration

If client is unable to provide sufficient data and if client approve, we can use above open dataset.

(v) Level of reliance on alternate data such as SMS in light of the fact that Android

ecosystem can bar the access to such data at any point.

Ans – Yes, all the above data can be easily access by any android ecosystem.

(vi) How would you avoid/minimize reinforcement of bias in your models?

Ans –

1. Selecting correct learning model.
2. Using correct representative dataset.
3. Monitoring performance using real data.

(vii) (Optional) Discuss elements of security infrastructure in the context of financial

services.