# *AI & ML*

# (SLOVED)MODEL QUESTION PAPER _ 02

NOTE : Answer one full question from each section:

## SECTION -1

**1A) With Industry 4.0, artificial intelligence is finding place in every aspect of life. What happens if AI replaces humans in the workplace?**

**ANS :** If AI replaces humans in the workplace, it could lead to job losses and economic disruption, but it could also lead to increased efficiency and productivity. It's important for society to consider how to mitigate the negative effects of AI on employment and to provide support for workers whose jobs are replaced by automation. Additionally, it's important to invest in education and training programs to help individuals acquire the skills necessary to work alongside AI and in industries that are less susceptible to automation.

**1B) For the given scenarios you are required to build an AI solution. Which AI techniques can be applied / best suited for stated problems. Justify**

1. **Extract and digitize the customer information from the Know Your Customer (KYC) forms.**

   **ANS**: Extracting and digitizing customer information from KYC forms: Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques can be used to extract and digitize the customer information from the forms. OCR can be used to extract text from an image and NLP can be used to extract structured data such as name, address, and ID number from the extracted text.

2. **To identify if employees are wearing face mask in the office campus**

   **ANS:** Identifying if employees are wearing face masks in the office campus: Computer Vision techniques can be used to identify if employees are wearing face masks in the office campus. A neural network-based image classifier can be trained using a dataset of images of people with and without masks, which can then be used to identify if employees are wearing masks in real-time.

3. **To identify and narrow down tumour regions and further predict if the tumour is malignant or not**

   **ANS:** Identifying and narrowing down tumour regions and further predict if the tumour is malignant or not: Medical imaging techniques such as Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI) can be used to identify and narrow down tumour regions. Machine learning techniques such as deep learning can be used to predict if the tumour is malignant or not. A deep learning model can be trained using a dataset of labelled medical images and then used to predict the malignancy of tumours.

4. **Automated inspection and cost estimation step in the Insurance claim business process**

   **ANS:** Automated inspection and cost estimation step in the Insurance claim business process: Computer Vision techniques can be used to automate the inspection step, where images of the damage are analysed to extract information about the extent of the damage. Natural Language Processing (NLP) techniques can be used to extract information from the textual description of the damage provided in the claim. In addition, a cost estimation algorithm can be used to estimate the cost of repairs based on the extracted information.

5. **To identify the location of a moving car within an image**

   **ANS:** Identifying the location of a moving car within an image: Computer vision and object detection techniques such as YOLO can be used to identify the location of a moving car within an image. A neural network-based object detector can be trained using a dataset of images of cars in various locations and then used to identify the location of a moving car in real-time.

**2A) Which technique help in addressing certain complex problems with higher accuracy and better generalization characteristics much like human brain in Computer Vision, Natural Language Processing and Speech Domains? And why?**

**ANS:** Deep learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been shown to be very effective in addressing complex problems in the computer vision, natural

language processing, and speech domains. The reason for this is that these types of neural networks are able to automatically learn hierarchical representations of the data, which is similar to how the human brain processes information. Additionally, they are able to handle large amounts of data and can be trained on multiple layers, which allows them to learn more complex patterns in the data.

**2B) For the following scenarios you are required to build a predictive model. Which machine learning technique/ algorithm can be applied / best suited for stated problems. Justify your recommendation.**

1. **Predicting the food delivery time**

   **ANS:** Predicting food delivery time: Regression algorithms such as linear regression or support vector regression would be well-suited for this task as the output is a continuous variable (delivery time) and the goal is to predict a numerical value.

2. **Predicting whether the transaction is fraudulent**

   **ANS:** Predicting whether a transaction is fraudulent: Classification algorithms such as logistic regression, decision tree, or random forest would be well-suited for this task as the output is a binary variable (fraud or not fraud) and the goal is to predict a class label.

3. **Predicting the credit limit of a credit card applicant**

   **ANS:** Predicting the credit limit of a credit card applicant: Regression algorithms such as linear regression or support vector regression would be well-suited for this task as the output is a continuous variable (credit limit) and the goal is to predict a numerical value.

4. **To group similar customers of an online grocery store, based on their purchasing patterns, to offer discounts to its customers.**

   **ANS:** Grouping similar customers of an online grocery store: Clustering algorithms such as k-means or hierarchical clustering would be well-suited for this task as the goal is to group similar customers based on their purchasing patterns.

5. **Predict the probability of a mechanical system breakdown, based on its system vibration and operating temperature**

   **ANS:** Predicting the probability of a mechanical system breakdown: Classification algorithms such as logistic regression, decision tree, or random forest would be well-suited for this task as the output is a binary variable (breakdown or not breakdown) and the goal is to predict a class label based on system vibration and operating temperature.

# SECTION -2

**3A) How to handle the missing values in the dataset? Explain.**

**ANS:** Handling missing values in a dataset is an important step in the pre-processing of the data before building a predictive model. There are several methods for handling missing values, and the appropriate method will depend on the specific dataset and the reason for the missing values. Some common methods include:

- Mean/Median/Mode Imputation: Replacing the missing value with the mean, median, or mode of the non-missing values in the same feature. This method is simple and easy to implement, but it can introduce bias if the missing values are not missing at random.

- Predictive Imputation: Using a predictive model to estimate the missing values based on the other features in the dataset. This method can be more accurate than mean/median/mode imputation, but it is also more computationally expensive.

- Multiple Imputation: Creating multiple imputed datasets by using a predictive model to estimate the missing values, and then combining the results to get a more accurate estimate.

- List-wise Deletion: Removing the entire observation from the dataset if it contains any missing values. This method can be useful when the missing values are missing completely at random.

- KNN Imputation: KNN imputation uses k-nearest neighbours to impute missing values. It computes the nearest neighbours for each observation with missing values, and then it uses the mean of the nearest neighbours as the imputed value.

- Using the Expectation-Maximization (EM) Algorithm : EM algorithm is a technique used to estimate the missing values. It is used to estimate the missing data in a dataset.

It's important to carefully consider the reason for the missing values and the potential impact on the results of the predictive model when choosing a method for handling missing values. Additionally, it's also important to keep track of the handling method for missing values, so that you can make sure that your model is not biased.

**3B) The statistical summary of Iris dataset is as follows.**

| | sepal-length | sepal-width | petal-length | petal-width |
|---|---|---|---|---|
| count | 150.000000 | 150.000000 | 150.000000 | 150.000000 |
| mean | 5.843333 | 3.054000 | 3.758667 | 1.198667 |
| std | 0.828066 | 0.433594 | 1.764420 | 0.763161 |
| min | 4.300000 | 2.000000 | 1.000000 | 0.100000 |
| 25% | 5.100000 | 2.800000 | 1.600000 | 0.300000 |
| 50% | 5.800000 | 3.000000 | 4.350000 | 1.300000 |
| 75% | 6.400000 | 3.300000 | 5.100000 | 1.800000 |
| max | 7.900000 | 4.400000 | 6.900000 | 2.500000 |

**Analyse and explain statistical metrics from above summary.**

**ANS:** The summary provided several statistical metrics for each of the four features (sepal-length, sepal-width, petal-length, and petal-width) of the Iris dataset. These metrics include:

- Count: The count is the number of observations in the dataset for each feature. In this case, all features have 150 observations.

- Mean: The mean is the average value of the observations for each feature. For example, the mean sepal-length is 5.8433333, meaning that the average sepal-length across all observations is 5.8433333 units.

- Standard Deviation (std): The standard deviation is a measure of the spread of the data around the mean. A low standard deviation indicates that the data points tend to be close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range.

- Minimum (min) and Maximum (max): These are the lowest and highest values of the observations for each feature. For example, the minimum sepal-length is 4.3 units and the maximum sepal-length is 7.9 units.

- Percentiles: These are values that indicate the point at which a certain percentage of the observations fall below. For example, 25% of the sepal-length observations fall below 5.1 units, 50% of the sepal-length observations fall below 5.8 units, and 75% of the sepal-length observations fall below 6.4 units.

It's important to note that the data provided in the summary seems to be not accurate and may be corrupted, as the values are not making sense. For example, sepal-width of 0 and 512.3292, petal-length of 1 and 6.9, petal-width of 0.1 and 2.5 are not realistic. This may lead to inaccurate predictions and results if a model is built based on this data. It's recommended to check the data for errors and missing values before proceeding with any analysis or modelling.

**4 A) Consider a real estate company that has a dataset containing the prices of properties in the Delhi region. It wishes to use the data to optimise the sale prices of the properties based on important factors such as area, bedrooms, parking, etc.**

**Essentially, the company wants —**

a. **To identify the variables affecting house prices, e.g. area, number of rooms, bathrooms, etc.**

b. **To create a model that quantitatively relates house prices with variables such as number of rooms, area, number of bathrooms, etc.**
c. **To know the accuracy of the model, i.e. how well these variables can predict house prices.**

**Discuss the steps to be followed to build such a model. Recommend the suitable techniques to consider at each step.**

**ANS:** To build a model to predict house prices based on factors such as area, bedrooms, parking, etc, the following steps can be followed:

1. Data Exploration and Cleaning:
   - First, the data must be explored to understand the variables available in the dataset and their distribution. This includes identifying missing values, outliers, and any other inconsistencies in the data.
   - The data must be cleaned and preprocessed to handle missing values, outliers, and any other inconsistencies.
   - It's recommended to use visualization techniques such as histograms, scatter plots, and box plots to explore the data and identify any patterns or relationships between the variables.

2. Feature Selection:
   - Next, the variables that are important for predicting house prices must be identified.
   - It's recommended to use techniques such as correlation analysis, chi-squared test, and mutual information to identify the most important variables.
   - It's also recommended to use techniques such as Recursive Feature Elimination (RFE) or SelectKBest to select the top features.

3. Model Building:
   - Once the important variables have been identified, a model can be built to predict house prices based on these variables.
   - For this problem, regression techniques are suitable. Linear regression, Random Forest Regression and Gradient Boosting Regression are the algorithms that can be considered.
   - It's recommended to use cross-validation to evaluate the performance of the model and select the best model.

4. Model Evaluation:
    - The model's accuracy can be evaluated using metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared.
    - It's also recommended to use techniques such as residual analysis, diagnostic plots, and model interpretability methods to understand the performance of the model and identify any potential issues.

5. Deployment:
    - Once the model is built and evaluated, it can be deployed in a production environment to make predictions on new data.
    - It's important to monitor the performance of the model and retrain it as needed to ensure it continues to make accurate predictions.
    - It's also important to note that to ensure the model is robust and generalizable to unseen data, it is important to have a good representation of data, which may require getting more data or doing more data pre-processing.

**4B) Describe univariate, bivariate, and multivariate analysis with suitable examples**

**ANS: Univariate Analysis:**

- Univariate analysis is the simplest form of statistical analysis that deals with one variable at a time. It is used to describe the basic features of the data in a variable such as the mean, median, mode, and standard deviation.

- For example, a retail store wants to analyse the sales of a particular product. They want to know the average sales, the highest and lowest sales, and the number of products sold during a particular time period. This can be done by performing univariate analysis on the sales data.

**Bivariate Analysis:**

- Bivariate analysis is the analysis of two variables at a time. It is used to understand the relationship between two variables and how one variable affects the other. The relationship can be visualized using graphs such as scatter plots, line plots, and bar charts.

- For example, a car manufacturer wants to analyze the relationship between the price of a car and the number of units sold. They can use bivariate analysis to create a scatter plot of the data, which will show the relationship between the price of the car and the number of units sold.

**Multivariate Analysis:**

- Multivariate analysis is the analysis of more than two variables at a time. It is used to understand the relationship between multiple variables and how they affect each other. It can be used to identify patterns and trends in the data.

- For example, an e-commerce company wants to analyze the relationship between the number of customer reviews, the customer rating, and the number of sales. They can use multivariate analysis to create a plot that shows the relationship between these three variables. This plot can help the company identify patterns and trends in the data, such as which products have the highest sales and which products have the best customer ratings.

It's worth noting that multivariate analysis can be a complex process that requires advanced statistical techniques such as Principal Component Analysis (PCA), Factor Analysis, and Cluster Analysis.

# SECTION -3

**5A) N-grams are defined as the combination of N keywords together. Consider the given sentence:**

**"Data Visualization is a way to express your data in a visual context so that patterns, correlations, trends between the data can be easily understood."**

**Generate bi-grams and tri-grams for the above sentence**

a. **Before performing text cleaning steps.**
   **ANS: Bi-grams**:

- "Data Visualization"
- "Visualization is"
- "is a"
- "a way"
- "way to"
- "to express"
- "express your"
- "your data"
- "data in"
- "in a"
- "a visual"
- "visual context"
- "context so"
- "so that"
- "that patterns"
- "patterns, correlations,"
- "correlations, trends"
- "trends between"
- "between the"
- "the data"
- "data can"
- "can be"
- "be easily"
- "easily understood."

**Tri-grams:**

- "Data Visualization is"
- "Visualization is a"
- "is a way"
- "a way to"
- "way to express"
- "to express your"
- "express your data"
- "your data in"
- "data in a"
- "in a visual"
- "a visual context"
- "visual context so"
- "context so that"
- "so that patterns"

- "that patterns, correlations,"
- "patterns, correlations, trends"
- "correlations, trends between"
- "trends between the"
- "between the data"
- "the data can"
- "data can be"
- "can be easily"
- "be easily understood."

**b. After performing following text cleaning steps:**

**i. Stop word Removal**

**ANS: Bi-grams:**

- "Data Visualization"
- "Visualization way"
- "way express"
- "express data"
- "data visual"
- "visual context"
- "context patterns"
- "patterns, correlations,"
- "correlations, trends"
- "trends data"
- "data easily"
- "easily understood."

**Tri-grams:**

- "Data Visualization way"
- "Visualization way express"
- "way express data"
- "express data visual"
- "data visual context"
- "visual context patterns"
- "context patterns, correlations,"
- "patterns, correlations, trends"
- "correlations, trends data"
- "trends data easily"
- "data easily understood."
- Note: Please note that by removing stop words, some of the bi-grams and tri-grams may be lost.

## ii. Replacing punctuations by a single space

**ANS:** After performing stop word removal and replacing punctuations by a single space, the sentence becomes:

"Data Visualization way express data visual context patterns correlations trends data easily understood"

The bi-grams for the cleaned sentence are:

- "Data Visualization"
- "Visualization way"
- "way express"
- "express data"
- "data visual"
- "visual context"
- "context patterns"
- "patterns correlations"
- "correlations trends"
- "trends data"
- "data easily"
- "easily understood"

The tri-grams for the cleaned sentence are:

- "Data Visualization way"
- "Visualization way express"
- "way express data"
- "express data visual"
- "data visual context"
- "visual context patterns"
- "context patterns correlations"
- "patterns correlations trends"
- "correlations trends data"
- "trends data easily"
- "data easily understood"

**5B) K-means clustering with Euclidean distance suffer from the curse of dimensionality. Is the statement true and why?**

- **ANS:** The statement is true. The curse of dimensionality refers to the phenomenon where the performance of many machine learning algorithms, including k-means clustering, decreases as the number of features (dimensions) in the data increases.

- One of the main reasons for this is that as the number of dimensions increases, the distance between any two points in the feature space becomes much larger, making it harder to find meaningful clusters. Additionally, as the number of dimensions increases, the number of data points required to accurately estimate the cluster centers also increases exponentially, making the algorithm computationally expensive.

- Euclidean distance, which is commonly used in k-means clustering, is sensitive to feature scaling, and with high dimensional data, the Euclidean distance is dominated by the dimensions with the largest scale, which leads to the algorithm to focus on those dimensions and it might fail to account for other dimensions.

- In summary, k-means clustering with Euclidean distance is sensitive to high dimensional data, and as the number of dimensions increases, the algorithm can become less effective and computationally expensive.

**6A) The sinking of the Titanic is one of the most infamous shipwrecks in history.On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew. You are asked to build a machine learning model to predict whether a passenger survived or not. Describe each step you will follow to build this model.**

**ANS:**

1. **Data Collection**: The first step would be to collect data on the passengers of the Titanic, including their demographic information

(such as age, gender, and socio-economic status), as well as information on their ticket and cabin class. This data can be obtained from various sources such as the Titanic dataset on Kaggle.

2. **Data Cleaning**: The next step would be to clean and preprocess the data, removing any missing or irrelevant information, and ensuring that the data is in a format that can be used for building a model. This may include handling missing values, converting categorical variables to numerical values, and standardizing the data.

3. **Feature Selection**: After cleaning the data, we would need to select the most relevant features for building the model. This could be done by using techniques such as correlation analysis, mutual information, and feature importance to identify the variables that are most important for predicting survival.

4. **Model Selection**: Once the data has been cleaned and features have been selected, we can select a suitable machine learning algorithm for building the model. Logistic regression, Random Forest, and SVM are some of the suitable algorithm which can be used for this problem.

5. **Model Training:** After selecting the appropriate algorithm, we would train the model using the selected features and the cleaned data. We would divide the data into training and testing sets, and use the training set to train the model.

6. **Model Evaluation:** After training the model, we would use the testing set to evaluate the model's performance. We would use metrics such as accuracy, precision, recall, and F1-score to evaluate the model's ability to predict survival.

7. **Model Optimization**: Based on the model's performance, we would optimize the model by adjusting the parameters or tuning the algorithm.

8. Deployment: Once the model has been optimized, we would deploy it in a production environment so that it can be used to predict the survival of passengers in real-world scenarios.

**6B) You work for a textile manufacturer and have been asked to build a model to detect and classify fabric defects. You trained a machine learning model with high recall. You want quality control inspectors to gain trust in your model. Which technique should you use to understand the rationale of your classifier? Justify**

**ANS:** One technique that can be used to understand the rationale of a classifier is feature importance analysis. This technique is used to determine which features or input variables the model is using to make predictions. By identifying which features are most important in the classification of fabric defects, quality control inspectors can gain insight into how the model is making its decisions and can therefore gain trust in the model.

One way to perform feature importance analysis is through the use of permutation importance. This is a technique that involves shuffling the values of a single feature and observing the effect on the model's predictions. By measuring the change in accuracy, we can determine the importance of that feature in the model's decision-making process. Another way is by using feature importance provided by tree based models like Random Forest, XGBoost etc.

Another technique which can be used to understand the rationale of classifier is interpretable models like Decision Trees and Rule-based models, etc. These models are easy to interpret and understand the decision making process of the model.

In summary, to understand the rationale of a classifier and help quality control inspectors gain trust in the model, feature importance analysis can be used to identify which features the model is using to make predictions. Additionally, interpretable models such as decision trees can also be used to make the model's decision-making process more transparent.

# SECTION -4

**7 A) A machine learning model was built to classify patient as covid +ve or -ve. The confusion matrix for the model is as shown below. Compute other performance metrics and analyse the performance of the model.**

| | | Actual | |
|---|---|---|---|
| | | 1 | 0 |
| Predicted | 1 | 397 | 103 |
| | 0 | 126 | 142 |

**ANS:**

The confusion matrix you provided shows the results of a binary classification model that was trained to predict whether a patient has COVID-19 (positive) or not (negative). The rows represent the actual values, and the columns represent the predicted values. The numbers in the matrix represent the number of observations that fall into each category.

From the confusion matrix, we can compute several performance metrics to evaluate the model's performance:

- **Accuracy**: (397 + 142) / (397 + 103 + 126 + 142) = 0.726 or 72.6%. This metric measures the proportion of correct predictions made by the model.
- **Precision**: 397 / (397 + 103) = 0.793 or 79.3%. This metric measures the proportion of true positive predictions among all positive predictions.
- **Recall (or Sensitivity or True Positive Rate):** 397 / (397 + 126) = 0.760 or 76.0%. This metric measures the proportion of actual positive observations that were correctly predicted as positive.
- **Specificity** : 142 / (142 + 103) = 0.580 or 58.0%. This metric measures the proportion of actual negative observations that were correctly predicted as negative
- **False Positive Rate (FPR) :** 103 / (142 + 103) = 0.420 or 42.0%. This metric measures the proportion of actual negative observations that were incorrectly predicted as positive

- **F1-Score:** 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.793 * 0.760) / (0.793 + 0.760) = 0.774 or 77.4%. This is the harmonic mean of precision and recall, which balances both measures.

It can be seen that the model has an accuracy of 72.6%. Which is not a good accuracy. The precision and recall are also not very good. The model is not classifying well.

**7B) A Machine Learning Engineer is preparing a data frame for a supervised learning task. The ML Engineer notices the target label classes are highly imbalanced and multiple feature columns contain missing values. The proportion of missing values across the entire data frame is less than 5%. What should the ML Engineer do to minimize bias due to missing values? Support your argument.**

**ANS:** The Machine Learning Engineer should take the following steps to minimize bias due to missing values:

1. **Imputation**: The ML Engineer should use a statistical imputation method, such as mean imputation or multiple imputation, to fill in the missing values. This approach would reduce the bias caused by the missing data by replacing missing values with plausible values based on the distribution of the observed data.
2. **Downsampling**: The ML Engineer should use the downsampling technique to reduce the imbalance of the target label classes. This would ensure that the model is not overfitting to one class while underfitting the other.
3. **Undersampling**: The ML Engineer should use the undersampling technique to reduce the imbalance of the target label classes. This would balance the number of observations in each class by randomly removing observations from the over-represented class.
4. **Upsampling**: The ML Engineer should use the upsampling technique to increase the number of observations in the under-represented class. This would balance the number of observations in each class by randomly replicating observations from the under-represented class.
5. **Combining**: The ML Engineer can also use a combination of the above-mentioned techniques for handling class imbalance. For example, using the combination of upsampling and downsampling.

By doing these steps, the ML Engineer can significantly reduce the bias caused by missing values and class imbalance in the data. Additionally, it is important to evaluate the performance of the model on an independent test dataset to ensure that the model generalizes well to new unseen data.

It is also important to mention that when we use imputation method, it is crucial to validate the assumption that missingness is missing at random (MAR). If the missing values are not missing at random, the imputed values will have a bias, and the model performance could be affected.

**8A) A data scientist is working on optimising a model during the training process by varying multiple parameters. The data scientist observes that, during multiple runs with the identical parameters the loss function converges to different, yet stable values. What should the data scientist do to improve the training process? Justify**

**ANS:** The data scientist should take the following steps to improve the training process:

1. Initialization: The data scientist should try different initialization methods for the model parameters such as random initialization, Xavier initialization, and He initialization. This can help to ensure that the model starts with a set of parameters that are more likely to converge to a good solution.
2. Regularization: The data scientist should try different types of regularization methods, such as L1 and L2 regularization, to prevent overfitting and improve the generalization of the model.
3. Data Augmentation: The data scientist should consider using data augmentation techniques to increase the size of the training dataset and reduce overfitting. This can help to improve the model's robustness to variations in the data.
4. Batch Normalization: The data scientist should consider using batch normalization to improve the training process by reducing the internal covariate shift.
5. Early Stopping: The data scientist should use early stopping to prevent overfitting by monitoring the performance of the model on a validation set and stopping the training process when the performance starts to degrade.

6. Different optimization Algorithm: The data scientist should consider changing the optimization algorithm being used. For example, try using Adam instead of SGD, or try different values of learning rate and momentum.

By taking these steps, the data scientist can improve the training process and reduce the variability in the final values of the loss function. Additionally, it is important to evaluate the performance of the model on an independent test dataset to ensure that the model generalizes well to new unseen data.

**8B) A company has collected customer comments on its products, rating them as safe or unsafe, using decision trees. The training dataset has the following features: id, date, full review, full review summary, and a binary safe/unsafe tag. During training, any data sample with missing features was dropped. In a few instances, the test set was found to be missing the full review text field. For this use case, which is the most effective course of action to address test data samples with missing features. Justify**

**ANS:** The most effective course of action to address test data samples with missing features would likely be to use a technique called imputation. Imputation is a method of estimating missing values in a dataset, which can be done using a variety of techniques such as mean imputation, median imputation, or multiple imputation. In this case, since the missing feature is the full review text field, using a technique such as multiple imputation could be appropriate, where a model is trained to predict the missing feature based on the other features in the dataset. This approach can help to minimize the potential loss of information and bias that can occur when simply dropping samples with missing features.

## SECTION -5

**9A) What are the deployment strategies borrowed from DevOps that can be utilized in MLOPs. Explain anyone strategy.**

**ANS:** There are several deployment strategies borrowed from DevOps that can be utilized in MLOps (Machine Learning Operations) to manage the deployment and maintenance of machine learning models in production. Some of these strategies include:

1. Continuous Integration and Continuous Deployment (CI/CD): This is a practice where code changes are automatically built, tested, and deployed to production, allowing for faster and more frequent updates to the model.
2. Blue-Green Deployment: This is a technique where two identical production environments are maintained, one "green" and one "blue." New updates are made to the "green" environment and tested, while the "blue" environment remains in production. Once the "green" environment is validated, the roles of the environments are switched, allowing for a seamless switchover with zero-downtime.
3. A/B Testing: This technique allows for multiple versions of the model to be deployed and tested simultaneously, with different subsets of users being directed to each version. This allows for easy comparison of the performance of different models and easy rollback to the previous version if needed.
4. Canary releases: This is a technique where a new version of the model is deployed to a small subset of users first, before being deployed to the entire user population. This allows for early detection of any issues that may arise before they affect the entire user base.

One of the strategies is Canary Release, In this strategy, a new version of the model is deployed to a small subset of users first, before being deployed to the entire user population. This allows for early detection of any issues that may arise before they affect the entire user base. This approach also allows for gradual roll-out of new features and models, which can be beneficial for testing the impact of new models on user engagement and satisfaction. Additionally, by monitoring the performance of the canary release, it is possible to identify and fix any issues before they can affect the entire user base.

**9B) Machine learning models can be resource heavy. They require a good amount of processing power to predict, validate, and recalibrate, millions of times over. How can containerisation of ML model solve this problem?**

**ANS:** Containerization of ML models can help to solve the problem of resource-intensive models by providing a lightweight and portable environment for deploying and running models. Containers are lightweight, standalone, executable packages that include everything needed to run a piece of software, including the code, runtime, system tools, libraries, and settings.

By containerizing an ML model, it can be deployed and run on any infrastructure that supports the container runtime, without requiring the underlying infrastructure to be configured specifically for the model. This makes it easier to scale the resources needed for the model and allows for more efficient use of resources, as multiple models can be run on the same infrastructure.

Additionally, containerization allows for better resource isolation and management, as each container runs in its own isolated environment with its own resources. This can help to prevent resource contention and ensure that each model has the resources it needs to perform well.

Another advantage of containerization is the ability to reproduce the same environment, which is critical for machine learning models. This allows for consistent results across different environments and makes it easier to troubleshoot issues.

Furthermore, containerization allows for easy orchestration and deployment using container orchestration tools like Kubernetes, Docker Swarm, etc. This makes it easier to manage the scaling and lifecycle of the containers, as well as automating the deployment of new versions of the model.

In summary, containerization allows for more efficient use of resources, better resource isolation, and management, consistent results, and easy orchestration and deployment of machine learning models.

**10A) How will you deploy a trained machine learning model as a predictive service in a production environment. Explain.**

**ANS:** Deploying a trained machine learning model as a predictive service in a production environment typically involves several steps, which can include the following:

1. Containerization: The model is packaged into a container using a tool such as Docker. This allows the model to be deployed and run in a lightweight and portable environment, making it easier to scale the resources needed for the model.
2. Model serving: The containerized model is deployed on a model serving framework such as TensorFlow Serving, Seldon Core, or Clipper. These frameworks provide an API for making predictions and managing the lifecycle of the model, including versioning and rollback.

3. Scaling: The predictive service is scaled to handle the expected traffic and load, by using a container orchestration tool like Kubernetes, Docker Swarm, etc. This allows for the automatic scaling of resources as needed, and makes it easier to manage the lifecycle of the service.
4. Monitoring: The predictive service is monitored for performance and errors, using tools such as Prometheus or Grafana for metrics and logging, and Sentry or Logstash for error tracking. This allows for early detection of any issues and helps to ensure that the service is performing well.
5. A/B Testing: A/B testing can be used to deploy multiple versions of the model and test them with different subsets of users. This allows for easy comparison of the performance of different models and easy rollback to the previous version if needed.
6. Deployment: The containerized model is deployed on a production environment, this can be on-premise, cloud-based or edge-based. This step can be automated using CI/CD pipeline to ensure that the deployment is consistent and easily repeatable.

In summary, deploying a machine learning model as a predictive service in a production environment involves containerizing the model, deploying it on a model serving framework, scaling the service to handle load, monitoring the service for performance and errors, and deploying the service in a production environment.

**10B) For the below given scenarios, suggest best suited cloud deployment model and list the challenges with it.**

**1. For ,**

**a. Variable workload**

**ANS:** For variable workload, the best suited cloud deployment model would be a hybrid cloud model. This allows for the workloads to be split between on-premise and cloud-based resources, giving greater flexibility and control over the resources used. The challenges with this model include the complexity of integrating on-premise and cloud-based resources, and the added cost of maintaining both environments.

**b. Test and Development**

**ANS**: For test and development, the best suited cloud deployment model would be a multi-cloud model. This allows for the use of different cloud

providers for different stages of the development process, giving greater flexibility and control over the resources used. The challenges with this model include the complexity of managing multiple cloud providers and the added cost of maintaining multiple environments.

**2. For,**

**a. Cloud bursting**

**ANS:** For cloud bursting, the best suited cloud deployment model would be a public cloud model. This allows for on-demand access to additional resources during peak usage times, reducing costs and increasing scalability. The challenges with this model include the potential for data security and privacy concerns and the need for a robust network infrastructure to handle the traffic between on-premise and cloud-based resources.

**b. On demand access**

**ANS:** For on demand access, the best suited cloud deployment model would be a public cloud model. This allows for easy scaling of resources as needed, and the ability to pay only for the resources used. The challenges with this model include the potential for data security and privacy concerns and the need for a robust network infrastructure to handle the traffic between on-premise and cloud-based resources.

**c. Sensitive data**

**ANS:** For sensitive data, the best suited cloud deployment model would be a private cloud model. This allows for greater control over the security and privacy of the data, and the ability to meet compliance requirements. The challenges with this model include the added cost of maintaining the private cloud environment, and the need for specialized expertise to manage and secure the environment.