



**RNS INSTITUTE OF TECHNOLOGY**

**Autonomous Institution Affiliated to VTU**

**Assignment 2: CLOUD COMPUTING**

## **University Question Bank**

*A compilation of important questions for review*

### **Question 1:**

#### **iii) Precision and Recall.?**

Answer: Precision and Recall Understanding Key Performance Metrics Precision and recall are crucial metrics used in information retrieval and machine learning, specifically in tasks involving classification or ranking. They help assess the performance of a model by measuring its ability to correctly identify relevant items and avoid misclassifying irrelevant items. Here's a breakdown

1. Precision Definition Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. Formula  $\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$  Interpretation A high precision value indicates that the model is accurate in its positive predictions, meaning it rarely predicts something as positive when it's actually negative. Example In spam detection, a high precision indicates that the model rarely flags legitimate emails as spam.

2. Recall Definition Recall measures the proportion of correctly predicted positive instances out of all actual positive instances. Formula  $\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$  Interpretation A high recall value indicates that the model is able to identify almost all relevant items, meaning it rarely misses positive instances. Example In disease diagnosis, a high recall means that the model is good at detecting the disease, even if it



**RNS INSTITUTE OF TECHNOLOGY**

**Autonomous Institution Affiliated to VTU**

## **Assignment 2: CLOUD COMPUTING**

might also flag some healthy individuals as having the disease. Relationship Precision and recall often have an inverse relationship. Increasing one may decrease the other. This is because a model that is more focused on catching all positive instances high recall might be less accurate in its predictions lower precision, and vice versa. Tradeoffs The ideal scenario is to have both high precision and high recall. However, this is not always achievable. The choice between emphasizing precision or recall depends on the specific application and the costs associated with false positives and false negatives. Example Imagine a model that predicts if a customer will click on an ad. High precision The model accurately predicts most clicks but might miss some. This is ideal if the cost of showing an irrelevant ad is high. High recall The model captures almost all potential clicks but might also include some false positives nonclicks. This is ideal if the cost of missing a potential click is high. F1Score To combine precision and recall into a single metric, the F1score is often used. It's the harmonic mean of precision and recall, giving equal weight to both. F1Score 2 Precision Recall Precision Recall Conclusion Understanding precision and recall is essential for evaluating the performance of classification and ranking models. They provide valuable insights into the model's ability to identify relevant items and avoid misclassifications. By considering both metrics and their tradeoffs, we can make informed decisions about model selection and optimization.

### **Question 2:**

#### **How to discover and visualize data to gain insights.?**

Answer: Discovering and Visualizing Data to Gain Insights A Comprehensive Guide This guide will walk you through the process of discovering and visualizing data to unearth meaningful insights. 1. Define Your Question What are you trying to achieve? Start with a clear objective.



## **RNS INSTITUTE OF TECHNOLOGY**

**Autonomous Institution Affiliated to VTU**

### **Assignment 2: CLOUD COMPUTING**

For example, "How does customer satisfaction relate to product usage?" What specific insights are you seeking? Do you want to understand trends, identify outliers, or compare different groups?

2. Gather Your Data Identify relevant data sources. This could include databases, spreadsheets, API data, social media feeds, or even user surveys. Ensure data quality. Check for missing values, inconsistencies, and errors. Data cleaning is crucial for reliable insights. Consider data types. Quantitative data numbers and qualitative data text, images require different analysis techniques.

3. Explore and Clean Your Data Understand your data structure. Explore relationships between variables, identify patterns, and understand data distribution. Perform basic data cleaning. This may involve handling missing values, correcting errors, and transforming data into a consistent format. Utilize data exploration tools. Data visualization tools like Tableau, Power BI, or even simple spreadsheets can help you gain initial insights.

4. Choose Appropriate Visualizations Select visualizations based on your data type and objective. Bar charts, line graphs, scatterplots, heatmaps, and treemaps are just a few examples. Use appropriate scales and labels. Ensure clarity and avoid misleading interpretations. Consider color and design. Choose colors and fonts that enhance readability and convey meaning effectively.

5. Analyze and Interpret Insights Focus on the story your data tells. Look for trends, outliers, correlations, and any other noteworthy patterns. Consider the context. Don't draw conclusions in isolation. Think about factors that may influence the data. Validate your insights. Compare your findings with other data sources or subject matter expertise.

6. Communicate Your Insights Create compelling visuals. Use clear, concise language and avoid technical jargon. Present your findings effectively. Consider your audience and choose an appropriate format reports, presentations, dashboards. Encourage interaction and discussion. Prompt questions and facilitate a dialogue around the data.

Tools and Resources Data Visualization Software



## **RNS INSTITUTE OF TECHNOLOGY**

**Autonomous Institution Affiliated to VTU**

### **Assignment 2: CLOUD COMPUTING**

Tableau, Power BI, Google Data Studio, Qlik Sense, Plotly Programming Languages Python with libraries like Pandas, Matplotlib, Seaborn, R, JavaScript OpenSource Visualization Libraries D3.js, VegaLite, Bokeh Key Considerations Data privacy and security Ensure you handle sensitive data responsibly. Ethical considerations Avoid misleading interpretations or drawing biased conclusions. Iteration and continuous improvement Data analysis is an iterative process. Regularly revisit your data and refine your analysis as needed. By following these steps and utilizing appropriate tools, you can effectively discover and visualize data to gain valuable insights and make informed decisions.

### **Question 3:**

#### **What is data cleaning?**

Answer: Data Cleaning Making Your Data Ready for Action Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and correcting errors, inconsistencies, and inaccuracies in your data. It's like giving your data a thorough makeover, removing clutter and flaws to make it more reliable and useful. Here's a breakdown Why is Data Cleaning Important? Accurate Insights Clean data leads to accurate analysis and meaningful insights. Imagine building a house on a shaky foundation that's what happens with bad data! Improved DecisionMaking Reliable data helps you make informed decisions that drive better results. Efficient Data Processing Clean data is easier to process, analyze, and store, leading to faster and smoother operations. Reduced Costs Addressing data issues early on saves you time and money in the long run, preventing errors and rework. Enhanced Data Integrity Ensures your data is consistent, reliable, and trustworthy. What are the Common Data Cleaning Tasks? Missing Values Filling in gaps in your data using methods like mean imputation or mode



## **RNS INSTITUTE OF TECHNOLOGY**

**Autonomous Institution Affiliated to VTU**

### **Assignment 2: CLOUD COMPUTING**

imputation. Outliers Identifying and correcting extreme values that don't fit the expected pattern. Inconsistent Data Standardizing data formats, units of measurement, and spelling. Duplicate Data Identifying and removing duplicate entries. Data Transformation Converting data into a suitable format for analysis. Data Validation Ensuring data conforms to predefined rules and standards. How is Data Cleaning Done? Manual Inspection Reviewing data visually and manually correcting errors. Automated Tools Using software programs to identify and fix data issues. Data Validation Rules Defining rules to check data quality and flag potential issues. Data Quality Metrics Evaluating data quality using metrics like accuracy, completeness, and consistency. Conclusion Data cleaning is a crucial step in any data analysis process. It ensures your data is accurate, reliable, and ready for use, leading to better insights, improved decisionmaking, and enhanced efficiency. By investing in data cleaning, you invest in the quality of your information and ultimately, the success of your projects.

#### **Question 4:**

**explain with suitable example.?**

Answer: Please provide me with the topic or concept you would like me to explain with an example. For example, you could ask Explain recursion with a suitable example. Explain the difference between a noun and a verb with a suitable example. Explain how to calculate the area of a triangle with a suitable example. Once you give me the topic, I will be happy to provide a clear explanation with a relevant example!