



WORD LIST PDF

Data Science Vocabulary List

Learn essential data science vocabulary with clear meanings, real-world examples, and related terms across data collection, data preparation, statistics, machine learning, model evaluation, analytics tools, visualization, and predictive modeling.

CATEGORY

Technology & Innovation

LEVEL

Intermediate

WORDS

86

Data science vocabulary is essential for understanding how data is collected, cleaned, analyzed, modeled, and used to solve real-world problems. From foundational terms like dataset, mean, median, variance, and correlation to advanced concepts such as feature engineering, cross-validation, deep learning, ROC curve, and predictive modeling, these words appear in statistics, machine learning, analytics, business intelligence, research, and modern technology discussions. This data science vocabulary list brings together carefully selected words related to data collection, data preparation, statistics, machine learning, model evaluation, data visualization, programming tools, and analytical workflows. Whether you are a beginner learning core terms like “data cleaning,” “SQL,” and “Pandas” or an advanced learner exploring concepts like “principal component analysis,” “bias-variance tradeoff,” “regularization,” and “ensemble learning,” this list helps you build strong, practical vocabulary for real-world use. Use this page to improve your reading, writing, speaking, and technical understanding of data science and analytics. Each word is designed to be clear, useful, and easy to connect with practical examples — helping students, English learners, analysts, researchers, and professionals communicate more confidently in data-driven environments.

OVERVIEW

Resource Summary

A quick overview of this LVF resource before you move into the lesson or vocabulary pages.

At a glance

- **Resource Type:** Word List PDF
- **Category:** Technology & Innovation
- **Level:** Intermediate
- **Total Words:** 86
- **Basic Words:** 17
- **Intermediate Words:** 23
- **Advanced Words:** 46
- **Brand:** LearnVocabularyFree.com

How to use this PDF

1. Read the definitions carefully.
2. Review examples and usage notes.
3. Try the practice section independently.
4. Check the answer key for self-review.

STARTER LIST

Words you will meet in this PDF (but not limited to)

- **Accuracy** — The proportion of correct predictions made by a model.
- **Association Rules** — Patterns that show relationships between items, often used in market basket analysis.
- **AUC (Area Under Curve)** — A metric that summarizes the overall performance of a model using the ROC curve.
- **Bayesian Statistics** — A statistical approach that updates probabilities as new evidence becomes available.
- **Bias-Variance Tradeoff** — The balance between model simplicity and model flexibility when trying to reduce prediction error.
- **Big Data** — Extremely large and complex datasets that require advanced tools and methods to store, process, and analyze.
- **Calculus** — A branch of mathematics focused on change and motion, used in optimization and machine learning.
- **Clustering** — An unsupervised learning technique used to group similar data points together.

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Accuracy

NOUN

BASIC

MEANING

The proportion of correct predictions made by a model.

EXAMPLE

The model achieved 92 percent accuracy on the test set.

SYNONYMS

correctness rate

Association Rules

NOUN

ADVANCED

MEANING

Patterns that show relationships between items, often used in market basket analysis.

EXAMPLE

Association Rules revealed which products were often bought together.

SYNONYMS

association analysis

AUC (Area Under Curve)

NOUN

ADVANCED

MEANING

A metric that summarizes the overall performance of a model using the ROC curve.

EXAMPLE

A higher AUC indicated better classification performance.

SYNONYMS

area under the ROC curve

Bayesian Statistics

NOUN

ADVANCED

MEANING

A statistical approach that updates probabilities as new evidence becomes available.

EXAMPLE

Bayesian statistics allows prior knowledge to be included in analysis.

SYNONYMS

Bayesian inference

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Bias-Variance Tradeoff

NOUN

ADVANCED

MEANING

The balance between model simplicity and model flexibility when trying to reduce prediction error.

EXAMPLE

The bias-variance tradeoff helped the team choose a better model complexity.

SYNONYMS

model complexity tradeoff

Big Data

NOUN

INTERMEDIATE

MEANING

Extremely large and complex datasets that require advanced tools and methods to store, process, and analyze.

EXAMPLE

Big Data technologies help companies process millions of transactions efficiently.

SYNONYMS

large-scale data

Calculus

NOUN

ADVANCED

MEANING

A branch of mathematics focused on change and motion, used in optimization and machine learning.

EXAMPLE

Calculus helps explain how gradient-based algorithms learn.

SYNONYMS

mathematical analysis

Clustering

NOUN

INTERMEDIATE

MEANING

An unsupervised learning technique used to group similar data points together.

EXAMPLE

Clustering helped the retailer identify different customer segments.

SYNONYMS

grouping analysis

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Computer Vision

NOUN

ADVANCED

MEANING

A field of AI and data science that enables computers to interpret images and visual information.

EXAMPLE

Computer Vision helps autonomous vehicles detect objects on the road.

SYNONYMS

image analysis

Confidence Interval

NOUN

ADVANCED

MEANING

A range of values that is likely to contain the true population value with a certain level of confidence.

EXAMPLE

The report presented a 95 percent confidence interval for the estimate.

SYNONYMS

interval estimate

Confusion Matrix

NOUN

ADVANCED

MEANING

A table used to evaluate classification results by comparing predicted and actual outcomes.

EXAMPLE

The confusion matrix showed where the model made mistakes.

SYNONYMS

classification matrix

Correlation

NOUN

INTERMEDIATE

MEANING

A measure of the strength and direction of the relationship between two variables.

EXAMPLE

There was a strong correlation between study time and exam scores.

SYNONYMS

association

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Covariance

NOUN

ADVANCED

MEANING

A measure of how two variables change together.

EXAMPLE

Positive covariance suggested both variables increased together.

SYNONYMS

joint variability

Cross-Validation

NOUN

ADVANCED

MEANING

A method of testing model performance by splitting data into multiple training and validation sets.

EXAMPLE

Cross-validation provided a more reliable estimate of model performance.

SYNONYMS

model validation

Data Cleaning

NOUN

BASIC

MEANING

The process of correcting errors, removing duplicates, and improving data quality.

EXAMPLE

Data cleaning removed duplicate customer records from the database.

SYNONYMS

data cleansing

data correction

Data Collection

NOUN

BASIC

MEANING

The process of gathering information from sources such as surveys, sensors, websites, or databases.

EXAMPLE

Data collection began with customer surveys and website activity logs.

SYNONYMS

information gathering

data gathering

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Data Lake

NOUN

ADVANCED

MEANING

A storage system that holds large amounts of raw structured and unstructured data.

EXAMPLE

The data lake stored logs, images, documents, and tabular files together.

SYNONYMS

raw data repository

Data Mining

NOUN

INTERMEDIATE

MEANING

The process of discovering useful patterns, trends, and relationships in large datasets.

EXAMPLE

Data mining helped the retailer identify customer buying patterns.

SYNONYMS

pattern discovery

knowledge discovery

Data Munging

NOUN

ADVANCED

MEANING

Another term for transforming and cleaning raw data so it becomes usable for analysis.

EXAMPLE

Data munging was necessary to combine files from multiple sources.

SYNONYMS

data wrangling

data preparation

Data Processing

NOUN

BASIC

MEANING

The process of organizing, transforming, and preparing raw data so it can be analyzed or used.

EXAMPLE

Data processing converted raw logs into a structured format for analysis.

SYNONYMS

data transformation

information processing

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Data Scraping

NOUN

INTERMEDIATE

MEANING

The automated extraction of data from websites or digital sources.

EXAMPLE

Data scraping was used to collect public pricing information from websites.

SYNONYMS

web scraping

data extraction

Data Scrubbing

NOUN

ADVANCED

MEANING

A process of finding and fixing inaccurate, incomplete, or inconsistent data.

EXAMPLE

Data scrubbing improved the reliability of the analytics report.

SYNONYMS

data cleaning

data cleansing

Data Visualization

NOUN

BASIC

MEANING

The presentation of data through charts, graphs, maps, and dashboards to make insights easier to understand.

EXAMPLE

Data visualization made the trends clear to non-technical stakeholders.

SYNONYMS

visual analytics

data graphics

Data Warehouse

NOUN

INTERMEDIATE

MEANING

A centralized system designed to store structured data for reporting and analysis.

EXAMPLE

The company used a data warehouse for historical business reporting.

SYNONYMS

analytics warehouse

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Data Wrangling

NOUN

INTERMEDIATE

MEANING

The process of cleaning, reshaping, and organizing raw data into a more useful format.

EXAMPLE

Data wrangling took most of the project time before analysis began.

SYNONYMS

data preparation

data reshaping

Dataset

NOUN

BASIC

MEANING

A structured collection of related data used for analysis, training, or reporting.

EXAMPLE

The dataset contained sales records from the last five years.

SYNONYMS

data collection

data file

Decision Tree

NOUN

INTERMEDIATE

MEANING

A machine learning model that uses a tree-like structure of decisions and rules to make predictions.

EXAMPLE

The decision tree classified loan applicants based on income and credit score.

SYNONYMS

tree model

Deep Learning

NOUN

ADVANCED

MEANING

A branch of machine learning that uses multi-layer neural networks to learn complex patterns from data.

EXAMPLE

Deep Learning improved the accuracy of the image classification system.

SYNONYMS

deep neural learning

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Descriptive Statistics

NOUN

INTERMEDIATE

MEANING

Statistical methods used to summarize and describe the main features of a dataset.

EXAMPLE

Descriptive statistics showed the average age and income of the sample.

SYNONYMS

summary statistics

ANTONYMS

inferential statistics

Dimensionality Reduction

NOUN

ADVANCED

MEANING

A technique used to reduce the number of features while preserving important information.

EXAMPLE

Dimensionality reduction helped simplify the dataset before visualization.

SYNONYMS

feature reduction

Ensemble Learning

NOUN

ADVANCED

MEANING

A machine learning approach that combines multiple models to improve prediction performance.

EXAMPLE

Ensemble Learning often performs better than a single model alone.

SYNONYMS

model combination

ETL (Extract, Transform, Load)

NOUN

INTERMEDIATE

MEANING

The process of extracting data from sources, transforming it, and loading it into a target system.

EXAMPLE

The ETL pipeline prepared daily sales data for reporting.

SYNONYMS

ETL pipeline

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Exploratory Data Analysis (EDA)

NOUN

INTERMEDIATE

MEANING

The process of examining data visually and statistically to understand patterns, trends, and issues before modeling.

EXAMPLE

Exploratory Data Analysis revealed missing values and strong correlations.

SYNONYMS

EDA

data exploration

F1 Score

NOUN

ADVANCED

MEANING

A metric that balances precision and recall into a single value.

EXAMPLE

The F1 Score helped compare models on an imbalanced dataset.

SYNONYMS

balanced metric

Feature Engineering

NOUN

ADVANCED

MEANING

The process of creating, transforming, or selecting variables to improve model performance.

EXAMPLE

Feature engineering helped the team improve prediction accuracy.

SYNONYMS

feature creation

variable transformation

Feature Selection

NOUN

ADVANCED

MEANING

The process of choosing the most relevant variables for a predictive model.

EXAMPLE

Feature selection reduced noise and improved efficiency.

SYNONYMS

variable selection

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Gradient Boosting (XGBoost / LightGBM)

NOUN

ADVANCED

MEANING

An ensemble learning method that builds models sequentially to correct previous errors and improve performance.

EXAMPLE

Gradient Boosting performed strongly in the competition dataset.

SYNONYMS

boosting

gradient boosting model

Hierarchical Clustering

NOUN

ADVANCED

MEANING

A clustering technique that builds nested groups of data points in a tree-like structure.

EXAMPLE

Hierarchical Clustering showed how the groups were related at different levels.

SYNONYMS

tree-based clustering

Hypothesis Testing

NOUN

ADVANCED

MEANING

A statistical method used to decide whether there is enough evidence to support a claim about a population.

EXAMPLE

Hypothesis testing was used to evaluate whether the new process improved results.

SYNONYMS

statistical testing

Inferential Statistics

NOUN

ADVANCED

MEANING

Statistical methods used to draw conclusions or make predictions about a population from sample data.

EXAMPLE

Inferential statistics helped the researcher generalize results from the sample.

SYNONYMS

statistical inference

ANTONYMS

descriptive statistics

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Jupyter Notebook

NOUN

BASIC

MEANING

An interactive environment used to write code, visualize data, and document analysis in one place.

EXAMPLE

The analyst shared the workflow through a Jupyter Notebook.

SYNONYMS

notebook environment

K-Means

NOUN

ADVANCED

MEANING

A clustering algorithm that groups data into a fixed number of clusters based on similarity.

EXAMPLE

K-Means divided the customers into four groups.

SYNONYMS

K-Means clustering

K-Nearest Neighbors (KNN)

NOUN

ADVANCED

MEANING

A machine learning algorithm that classifies or predicts values based on the closest data points.

EXAMPLE

K-Nearest Neighbors labeled the new data point by comparing nearby examples.

SYNONYMS

KNN

Linear Algebra

NOUN

ADVANCED

MEANING

A branch of mathematics dealing with vectors, matrices, and linear equations, widely used in data science and machine learning.

EXAMPLE

Linear algebra is essential for understanding many machine learning algorithms.

SYNONYMS

matrix mathematics

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Linear Regression

NOUN

INTERMEDIATE

MEANING

A supervised learning method used to predict a continuous value based on the relationship between variables.

EXAMPLE

Linear regression was used to estimate housing prices.

SYNONYMS

regression model

Logistic Regression

NOUN

ADVANCED

MEANING

A supervised learning algorithm used for classification problems, especially binary outcomes.

EXAMPLE

Logistic regression predicted whether customers would churn or stay.

SYNONYMS

classification regression

Machine Learning (Supervised)

NOUN

INTERMEDIATE

MEANING

A type of machine learning where models are trained on labeled data with known outcomes.

EXAMPLE

Machine Learning (Supervised) is used in spam detection and price prediction.

SYNONYMS

supervised learning

ANTONYMS

unsupervised learning

Mean

NOUN

BASIC

MEANING

The arithmetic average of a set of values.

EXAMPLE

The mean test score was 72.

SYNONYMS

average

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Mean Squared Error (MSE)

NOUN

ADVANCED

MEANING

A metric that measures the average squared difference between predicted and actual values.

EXAMPLE

The model with the lowest MSE was selected.

SYNONYMS

MSE

Median

NOUN

BASIC

MEANING

The middle value in a sorted list of numbers.

EXAMPLE

The median income was lower than the mean because of a few very high values.

SYNONYMS

middle value

Missing Values

NOUN

BASIC

MEANING

Data points that are absent or not recorded in a dataset.

EXAMPLE

The analyst handled missing values before training the model.

SYNONYMS

missing data

null values

Mode

NOUN

BASIC

MEANING

The value that appears most frequently in a dataset.

EXAMPLE

The mode of the shoe sizes was 42.

SYNONYMS

most frequent value

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Natural Language Processing (NLP)

NOUN

ADVANCED

MEANING

A field of AI and data science focused on helping computers understand, analyze, and generate human language.

EXAMPLE

Natural Language Processing is used in translation and chatbots.

SYNONYMS

NLP

Neural Networks

NOUN

ADVANCED

MEANING

Computational models inspired by the brain that learn patterns from data using interconnected layers.

EXAMPLE

Neural Networks are widely used in image recognition and language tasks.

SYNONYMS

artificial neural networks

Normalization

NOUN

ADVANCED

MEANING

The process of scaling data into a standard range, often between 0 and 1.

EXAMPLE

Normalization helped prepare the features for machine learning.

SYNONYMS

rescaling

feature scaling

NoSQL

NOUN

INTERMEDIATE

MEANING

A category of databases designed for flexible, non-relational data storage and large-scale systems.

EXAMPLE

NoSQL databases are useful for unstructured or fast-changing data.

SYNONYMS

non-relational database

ANTONYMS

SQL

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Null Hypothesis

NOUN

ADVANCED

MEANING

A default assumption that there is no effect, no difference, or no relationship between variables.

EXAMPLE

The null hypothesis stated that the new treatment had no impact.

SYNONYMS

baseline hypothesis

ANTONYMS

alternative hypothesis

NumPy

NOUN

BASIC

MEANING

A Python library used for numerical computing, arrays, and mathematical operations.

EXAMPLE

NumPy improved performance when working with large numeric arrays.

SYNONYMS

numerical Python library

Outlier

NOUN

INTERMEDIATE

MEANING

A data point that is very different from the rest of the observations in a dataset.

EXAMPLE

The unusual income value was treated as an outlier.

SYNONYMS

anomaly

extreme value

Overfitting

NOUN

INTERMEDIATE

MEANING

A modeling problem where a model learns the training data too closely and performs poorly on new data.

EXAMPLE

The model showed overfitting because it failed on the validation set.

SYNONYMS

poor generalization

ANTONYMS

good generalization

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

P-value

NOUN

ADVANCED

MEANING

A statistical measure used to judge the strength of evidence against the null hypothesis.

EXAMPLE

A low p-value suggested that the result was statistically significant.

SYNONYMS

probability value

Pandas

NOUN

BASIC

MEANING

A Python library used for data manipulation and analysis with tables and time-series data.

EXAMPLE

Pandas made it easy to filter, group, and summarize the dataset.

SYNONYMS

Python data library

Power BI

NOUN

BASIC

MEANING

A business intelligence and data visualization tool used to create reports and dashboards.

EXAMPLE

Power BI was used to build a sales dashboard for executives.

SYNONYMS

business intelligence tool

Precision

NOUN

ADVANCED

MEANING

The proportion of positive predictions that are actually correct.

EXAMPLE

Precision matters when false positives are costly.

SYNONYMS

positive prediction quality

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Predictive Modeling

NOUN

INTERMEDIATE

MEANING

The use of data and algorithms to forecast future outcomes or behaviors.

EXAMPLE

Predictive modeling helped estimate which customers were likely to leave.

SYNONYMS

prediction modeling

forecast modeling

Principal Component Analysis (PCA)

NOUN

ADVANCED

MEANING

A dimensionality reduction technique that transforms variables into a smaller set of principal components.

EXAMPLE

PCA was used to reduce feature complexity before clustering.

SYNONYMS

PCA

Probability Distribution

NOUN

ADVANCED

MEANING

A mathematical description of how values are likely to be spread across possible outcomes.

EXAMPLE

The normal distribution is a common probability distribution in statistics.

SYNONYMS

distribution

Python / R

NOUN

BASIC

MEANING

Two popular programming languages widely used in data science, machine learning, and statistical analysis.

EXAMPLE

Python and R are both powerful tools for data science projects.

SYNONYMS

data science programming languages

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

PyTorch

NOUN

INTERMEDIATE

MEANING

An open-source deep learning framework popular in research and production AI systems.

EXAMPLE

PyTorch is often preferred for flexible experimentation in deep learning.

SYNONYMS

deep learning framework

R-squared

NOUN

ADVANCED

MEANING

A statistical measure showing how much of the variation in the target variable is explained by a model.

EXAMPLE

The model had an R-squared of 0.81, indicating strong explanatory power.

SYNONYMS

coefficient of determination

Random Forest

NOUN

ADVANCED

MEANING

A machine learning method that combines many decision trees to improve prediction accuracy and stability.

EXAMPLE

Random forest performed better than a single decision tree on the test set.

SYNONYMS

ensemble of trees

Recall

NOUN

ADVANCED

MEANING

The proportion of actual positive cases that a model correctly identifies.

EXAMPLE

High recall is important in disease detection systems.

SYNONYMS

sensitivity

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Recommendation System

NOUN

ADVANCED

MEANING

A system that suggests products, content, or actions based on user data and patterns.

EXAMPLE

The streaming platform uses a recommendation system to suggest movies.

SYNONYMS

recommender system

Regularization (L1 / L2)

NOUN

ADVANCED

MEANING

A technique used to reduce overfitting by adding a penalty to model complexity.

EXAMPLE

Regularization improved performance by limiting unnecessary feature influence.

SYNONYMS

penalty method

Reinforcement Learning

NOUN

ADVANCED

MEANING

A type of machine learning where an agent learns through trial and error by receiving rewards or penalties.

EXAMPLE

Reinforcement Learning is often used in robotics and game-playing systems.

SYNONYMS

reward-based learning

ROC Curve

NOUN

ADVANCED

MEANING

A graph used to evaluate classification performance across different threshold settings.

EXAMPLE

The ROC Curve helped compare the classifier's sensitivity and specificity.

SYNONYMS

receiver operating characteristic curve

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Root Mean Squared Error (RMSE)

NOUN

ADVANCED

MEANING

A metric that measures the typical size of prediction errors in the same units as the target variable.

EXAMPLE

RMSE made it easier to interpret model error in practical terms.

SYNONYMS

RMSE

Scikit-learn

NOUN

INTERMEDIATE

MEANING

A Python machine learning library used for classification, regression, clustering, and evaluation.

EXAMPLE

Scikit-learn was used to train and test several machine learning models.

SYNONYMS

sklearn

SQL (Structured Query Language)

NOUN

BASIC

MEANING

A language used to store, manage, and query data in relational databases.

EXAMPLE

SQL was used to extract customer records from the warehouse.

SYNONYMS

SQL

Standard Deviation

NOUN

INTERMEDIATE

MEANING

A measure of how much the values in a dataset typically differ from the mean.

EXAMPLE

A low standard deviation means the values are clustered closely together.

SYNONYMS

dispersion measure

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Standardization

NOUN

ADVANCED

MEANING

The process of transforming data so it has a mean of 0 and a standard deviation of 1.

EXAMPLE

Standardization made the variables easier to compare in the model.

SYNONYMS

z-score scaling

Support Vector Machine (SVM)

NOUN

ADVANCED

MEANING

A supervised learning algorithm used for classification and regression by finding the best boundary between groups.

EXAMPLE

The Support Vector Machine separated the categories with a clear margin.

SYNONYMS

SVM

Tableau

NOUN

BASIC

MEANING

A data visualization tool used to create interactive charts, dashboards, and reports.

EXAMPLE

The team used Tableau to present business insights to management.

SYNONYMS

visual analytics tool

TensorFlow

NOUN

INTERMEDIATE

MEANING

An open-source framework widely used for machine learning and deep learning applications.

EXAMPLE

TensorFlow helped the team build a neural network for image recognition.

SYNONYMS

ML framework

Data Science Vocabulary List

Definitions, examples, and related vocabulary prepared in a structured LVF format.

Underfitting

NOUN

INTERMEDIATE

MEANING

A modeling problem where a model is too simple to capture important patterns in the data.

EXAMPLE

Underfitting caused the model to perform badly on both training and test data.

SYNONYMS

insufficient learning

ANTONYMS

good fit

Variance

NOUN

INTERMEDIATE

MEANING

A measure of how far values in a dataset are spread out from the mean.

EXAMPLE

High variance indicated that the data points were widely dispersed.

SYNONYMS

spread measure



PRACTICE

Mini Worksheet

Use the meanings and context from this PDF to answer the following practice questions.

1. Write the meaning of *Accuracy* and use it in one sentence.

.....
.....
.....

2. Write the meaning of *Association Rules* and use it in one sentence.

.....
.....
.....

3. Write the meaning of *AUC (Area Under Curve)* and use it in one sentence.

.....
.....
.....

4. Write the meaning of *Bayesian Statistics* and use it in one sentence.

.....
.....
.....

5. Write the meaning of *Bias-Variance Tradeoff* and use it in one sentence.

.....
.....
.....

6. Write the meaning of *Big Data* and use it in one sentence.

.....
.....
.....

ANSWER KEY

Suggested Answers

These answer points are designed to support self-review and classroom guidance.

1. Accuracy

Meaning:

The proportion of correct predictions made by a model.

Example:

The model achieved 92 percent accuracy on the test set.

2. Association Rules

Meaning:

Patterns that show relationships between items, often used in market basket analysis.

Example:

Association Rules revealed which products were often bought together.

3. AUC (Area Under Curve)

Meaning:

A metric that summarizes the overall performance of a model using the ROC curve.

Example:

A higher AUC indicated better classification performance.

4. Bayesian Statistics

Meaning:

A statistical approach that updates probabilities as new evidence becomes available.

Example:

Bayesian statistics allows prior knowledge to be included in analysis.

5. Bias-Variance Tradeoff

Meaning:

The balance between model simplicity and model flexibility when trying to reduce prediction error.

Example:

The bias-variance tradeoff helped the team choose a better model complexity.

6. Big Data

Meaning:

Extremely large and complex datasets that require advanced tools and methods to store, process, and analyze.

Example:

Big Data technologies help companies process millions of transactions efficiently.