AKTU · 3RD YEAR · SEMESTER 5 · BCDS051

Business Intelligence and Analytics

Unit 4 — BI – Data Analytics Process

- Introduction to Analytics Process
- Types of Analytical Techniques in BI – Descriptive, Predictive, Prescriptive
- Social Media Analytics
- Behavioural Analytics
- Iris Datasets



DECODED BY

Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

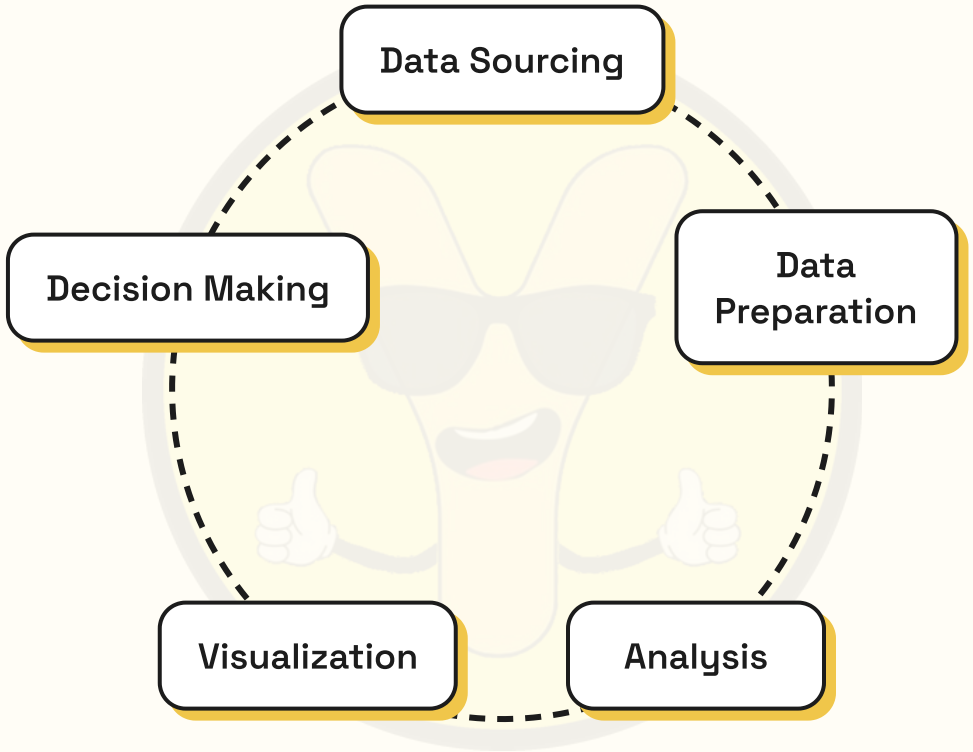
theycotes.com

Introduction to Analytics Process

The analytics process is a systematic approach to transforming raw data into actionable business insights.

- Data Collection: Gathering data from diverse sources like ERP, CRM, and external APIs.
- Data Cleaning: Removing noise, handling missing values, and correcting inconsistencies.
- Data Integration: Combining data from different sources into a unified view.
- Data Analysis: Applying statistical and mathematical models to identify patterns.
- Interpretation: Translating technical results into business-relevant recommendations.
- Action: Implementing decisions based on the derived insights.

Analytics Process Cycle

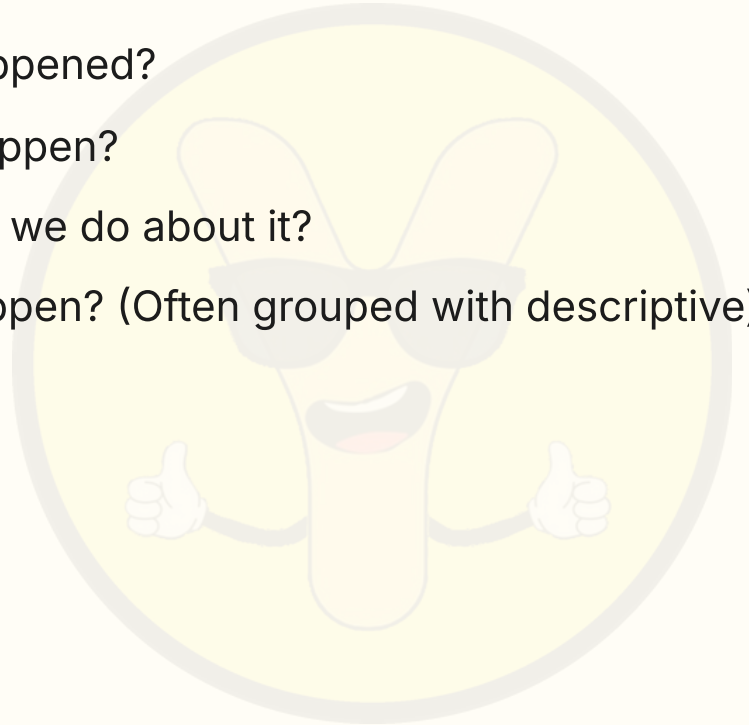


The iterative cycle of the data analytics process.

Types of Analytical Techniques

BI analytics is categorized based on the complexity of the question being answered.

- Descriptive Analytics: What has happened?
- Predictive Analytics: What could happen?
- Prescriptive Analytics: What should we do about it?
- Diagnostic Analytics: Why did it happen? (Often grouped with descriptive).

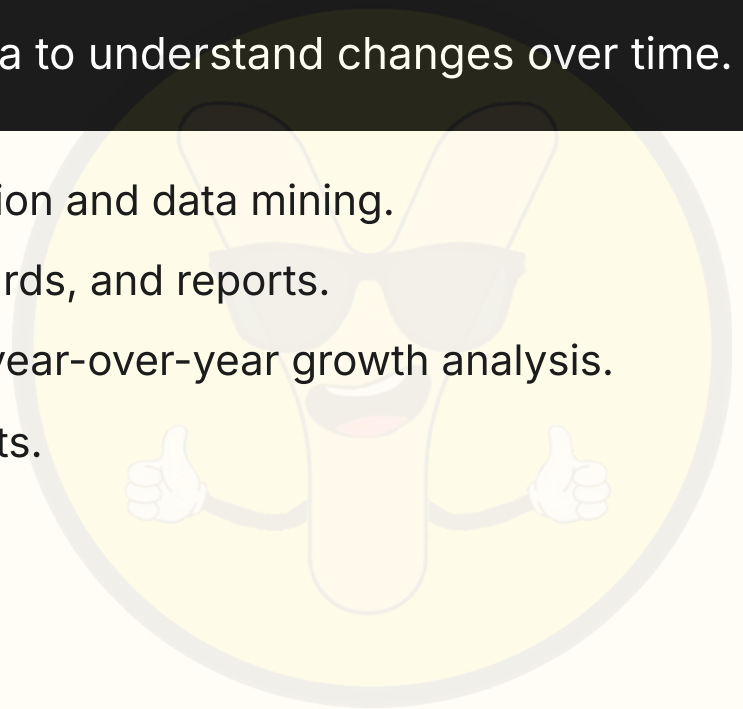


Descriptive Analytics

DESCRIPTIVE ANALYTICS

The interpretation of historical data to understand changes over time.

- Uses techniques like data aggregation and data mining.
- Primary tools: Dashboards, scorecards, and reports.
- Example: Monthly sales reports or year-over-year growth analysis.
- Focuses on summarizing past events.



Predictive Analytics

PREDICTIVE ANALYTICS

The use of statistical models and forecasting techniques to understand the future.

- Relies on historical data to identify trends and probabilities.
- Common algorithms: Linear Regression, Logistic Regression, Decision Trees.
- Example: Predicting customer churn or future stock prices.
- Outcome is a probability score, not a certainty.

Prescriptive Analytics

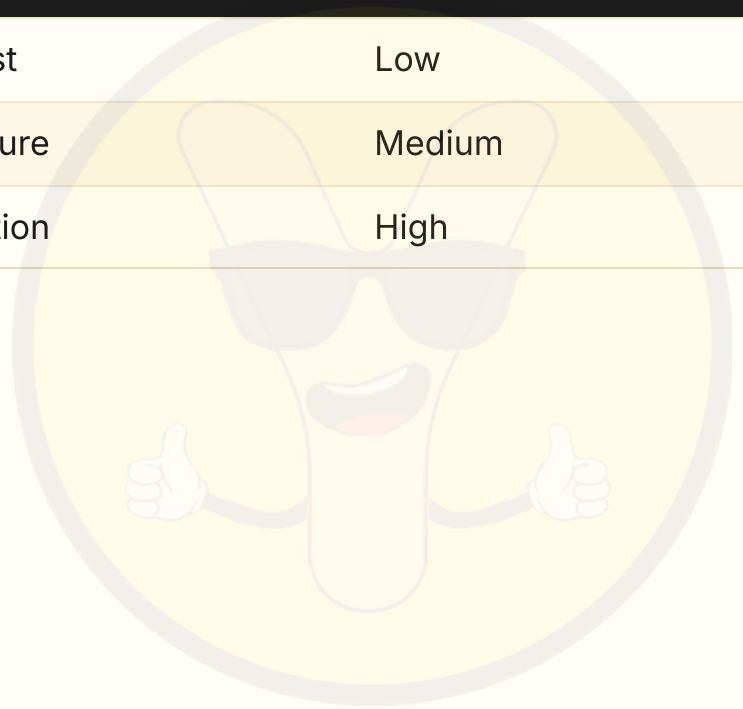
PRESCRIPTIVE ANALYTICS

The highest level of analytics that suggests specific actions to achieve desired outcomes.

- Uses optimization and simulation algorithms.
- Answers the question: How can we make it happen?
- Example: Dynamic pricing models or supply chain optimization.
- Requires complex mathematical modeling and real-time data.

Comparison of Analytics Types

Type	Focus	Complexity	Value
Descriptive	Past	Low	Low
Predictive	Future	Medium	Medium
Prescriptive	Action	High	High



Social Media Analytics

The process of gathering and analyzing data from social media platforms to drive business decisions.

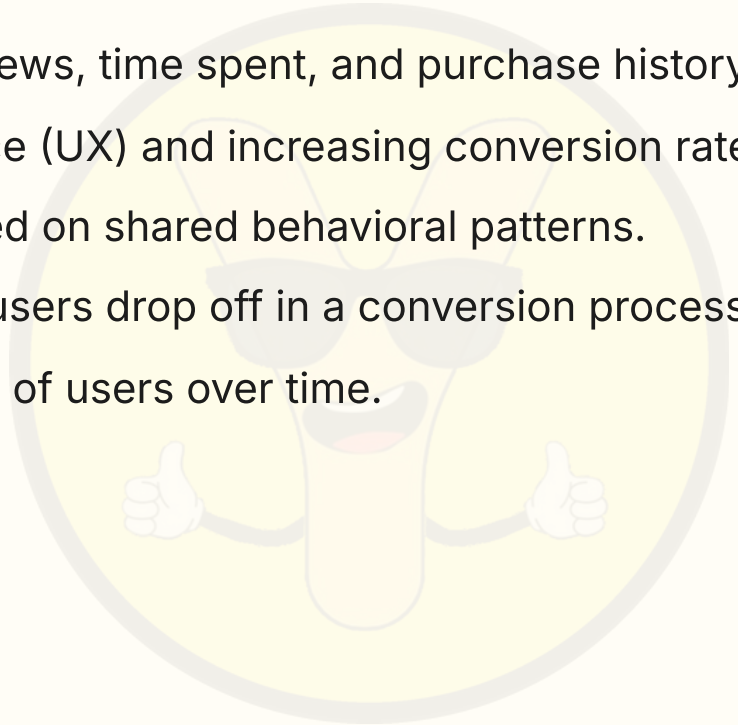
- Sentiment Analysis: Determining if public opinion is positive, negative, or neutral.
- Reach and Engagement: Tracking likes, shares, comments, and follower growth.
- Brand Monitoring: Tracking mentions of the brand across the web.
- Influencer Identification: Finding key individuals who drive brand conversations.
- Data Sources: Twitter API, Facebook Insights, LinkedIn Analytics.



Behavioural Analytics

Behavioural analytics focuses on understanding how users interact with products or services.

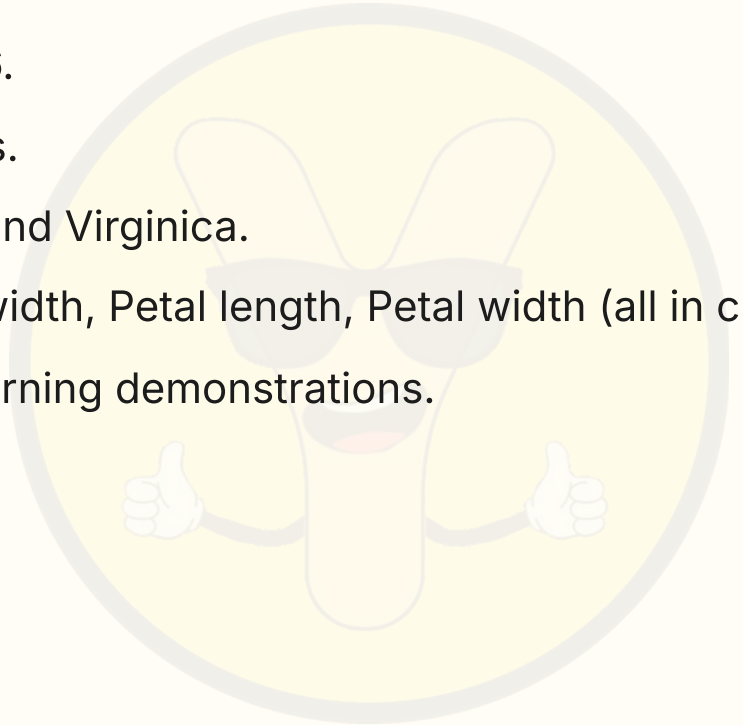
- Tracks user actions: Clicks, page views, time spent, and purchase history.
- Goal: Optimizing the user experience (UX) and increasing conversion rates.
- Segmentation: Grouping users based on shared behavioral patterns.
- Funnel Analysis: Identifying where users drop off in a conversion process.
- Cohort Analysis: Comparing groups of users over time.



Iris Dataset Overview

The Iris dataset is a classic multivariate dataset used for testing classification algorithms.

- Introduced by Ronald Fisher in 1936.
- Contains 150 samples of iris flowers.
- Three species: Setosa, Versicolor, and Virginica.
- Four features: Sepal length, Sepal width, Petal length, Petal width (all in cm).
- Used extensively for supervised learning demonstrations.

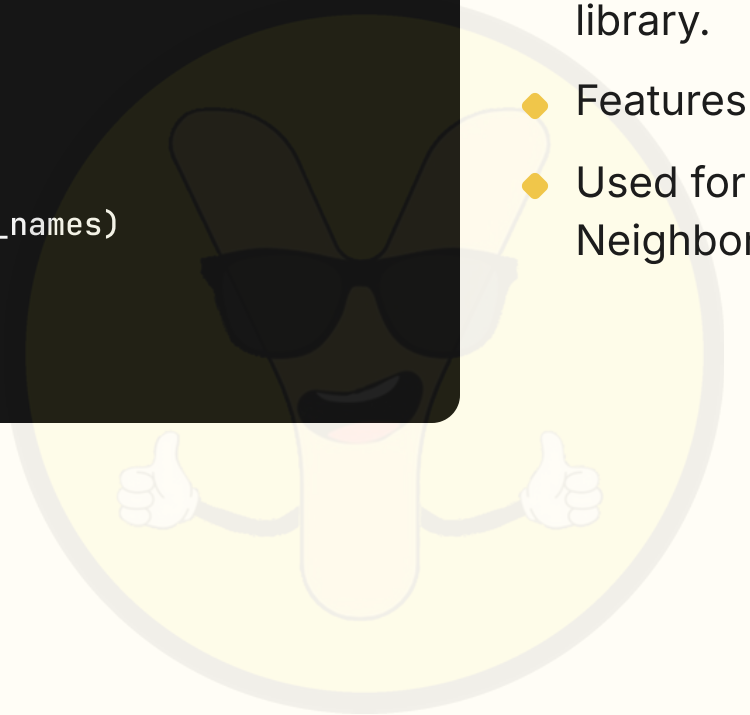


Iris Dataset Structure (Python)

```
from sklearn.datasets import load_iris
import pandas as pd

iris = load_iris()
df = pd.DataFrame(iris.data,
                  columns=iris.feature_names)
df['species'] = iris.target

print(df.head())
```



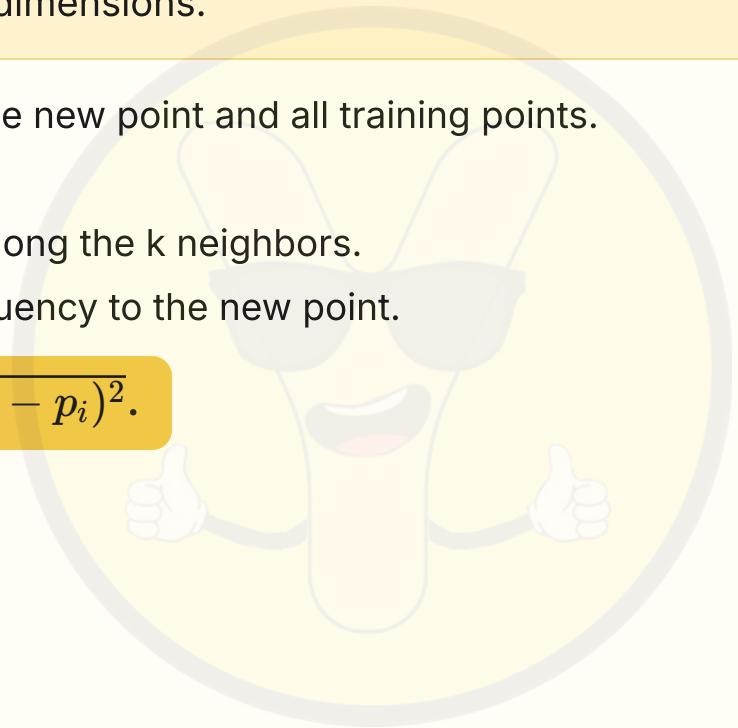
- The dataset is included in the Scikit-learn library.
- Features are stored in 'data', labels in 'target'.
- Used for training models like K-Nearest Neighbors.

K-Nearest Neighbors (KNN) on Iris

Classify an iris flower given sepal/petal dimensions.

1. Calculate Euclidean distance between the new point and all training points.
2. Find the 'k' closest points (neighbors).
3. Count the frequency of each species among the k neighbors.
4. Assign the species with the highest frequency to the new point.

The distance formula is $d = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$.



Exam corner

- **Q1** Explain the stages of the Data Analytics Process with a neat diagram.
- **Q2** Differentiate between Descriptive, Predictive, and Prescriptive analytics.
- **Q3** How does Social Media Analytics help in brand reputation management?
- **Q4** Define Behavioural Analytics and discuss its importance in e-commerce.
- **Q5** Describe the features and significance of the Iris dataset in machine learning.
- **Q6** Explain how predictive analytics uses historical data to forecast future trends.
- **Q7** Why is Prescriptive analytics considered the most advanced form of BI?

Unit 4 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 5 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

