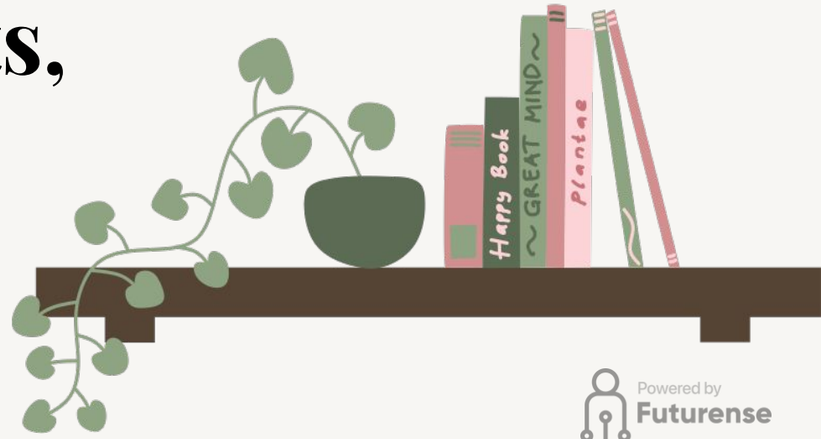
BS/BSc

Applied AI & Data Science

Predicting Employee Attrition Using Weka, Google Sheets, Flourish & Napkin AI



Powered by Futurense

Mini Project Session Plan (PRP)

Predicting Employee Attrition Using Weka, Google Sheets, Flourish & Napkin AI

A complete end-to-end ML workflow for beginners

Powered by Futurense

Agenda

01 Corporate Story

02 Dataset Overview

03 Dataset Generation Prompt

04 Public Dataset Option

05 Mini-Project Overview

06 Step-by-Step Execution Guide

07 Overfitting Activity

08 Underfitting Activity

09 Clues & Hints

10 Expected Learning Outcomes

11 Submission Checklist

Powered by Futurense

Corporate Story: Why This Matters

TechNova Solutions faces increasing employee attrition. HR, Managers, Finance & Leadership are affected.

Decisions impacted:

- Salary hikes
- Department-level interventions
- Predicting who will leave
- Training & upskilling

Data analysis + ML models help anticipate attrition and improve retention strategies.

Powered by Futurense

How Tools Solve This Problem

Google Sheets

Clean and prepare data

Weka

Build predictive models

Flourish

Visual dashboards and insights

Napkin AI

Hypothesis building & interpretation

Together, these simulate real HR Analytics workflows used in companies like TCS, Accenture, Infosys, Amazon.

Powered by Futurense

Dataset Specification (1/2)

Dataset Name:

TechNova Employee Attrition Dataset

Rows:

500

Columns:

15

Target Variable:

Attrition (Yes/No)

Features include:

- EmployeeID
- Age
- Gender
- Department
- JobRole
- MonthlyIncome
- YearsAtCompany
- Promotions
- JobSatisfaction
- WorkLifeBalance
- EducationLevel
- TrainingHours
- OverTime
- PerformanceRating

Powered by Futurense

Dataset Specification (2/2)

Why this dataset works:

Mix of numeric + categorical fields

Strong correlation patterns for ML

Perfect for classification tasks in Weka

Helps demonstrate both overfitting and underfitting

Powered by Futurense

Synthetic Dataset Generation Prompt

Copy and paste this into ChatGPT:

```
Generate a synthetic CSV dataset with 500 rows for an HR attrition prediction problem.Columns required:EmployeeID (1001–1500)Age (22–58)Gender (Male/Female)Department (IT, Sales, HR, Finance, R&D)JobRole (realistic roles)MonthlyIncome (15000–150000)YearsAtCompany (0–20)YearsSinceLastPromotion (0–10)JobSatisfaction (1–4)WorkLifeBalance (1–4)EducationLevel (1–5)TrainingHoursLastYear (0–80)OverTime (Yes/No)PerformanceRating (1–5)Attrition (target)Ensure correlations & no missing values.
```

Powered by Futurense

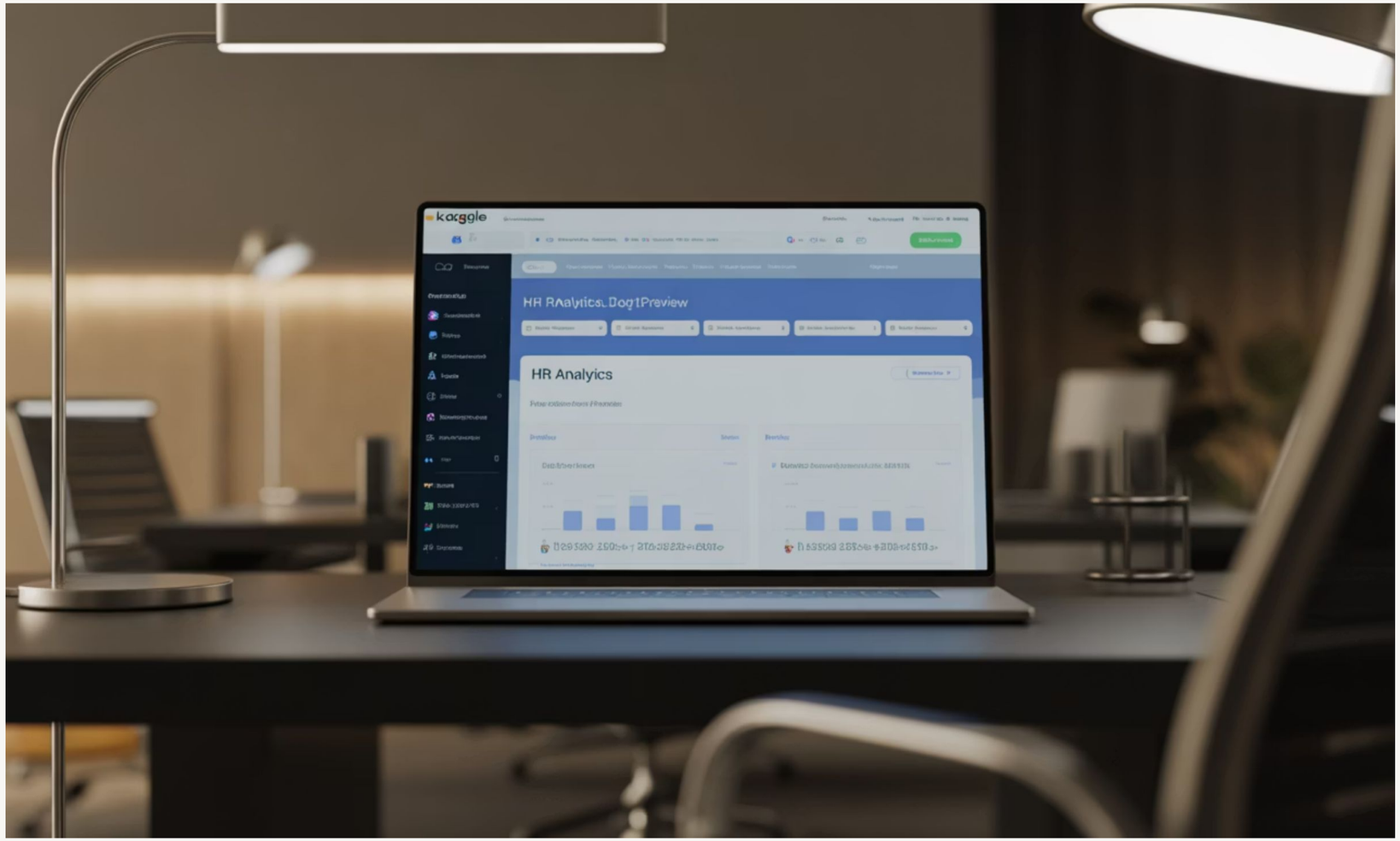
Public Dataset Recommendation

IBM HR Analytics Attrition Dataset

Link: Kaggle → IBM HR Analytics

Why suitable:

- Realistic HR data
- Industry-standard for attrition prediction
- Easy to load in Weka



Powered by Futurense

Mini Project Overview

Students will:

- Clean dataset in Google Sheets
- Visualize patterns in Flourish
- Train ML models in Weka
- Detect overfitting & underfitting
- Interpret model outcomes using Napkin AI
- Submit all screenshots in a single PDF

Powered by Futurense

Step 1: Load Dataset

Tool:

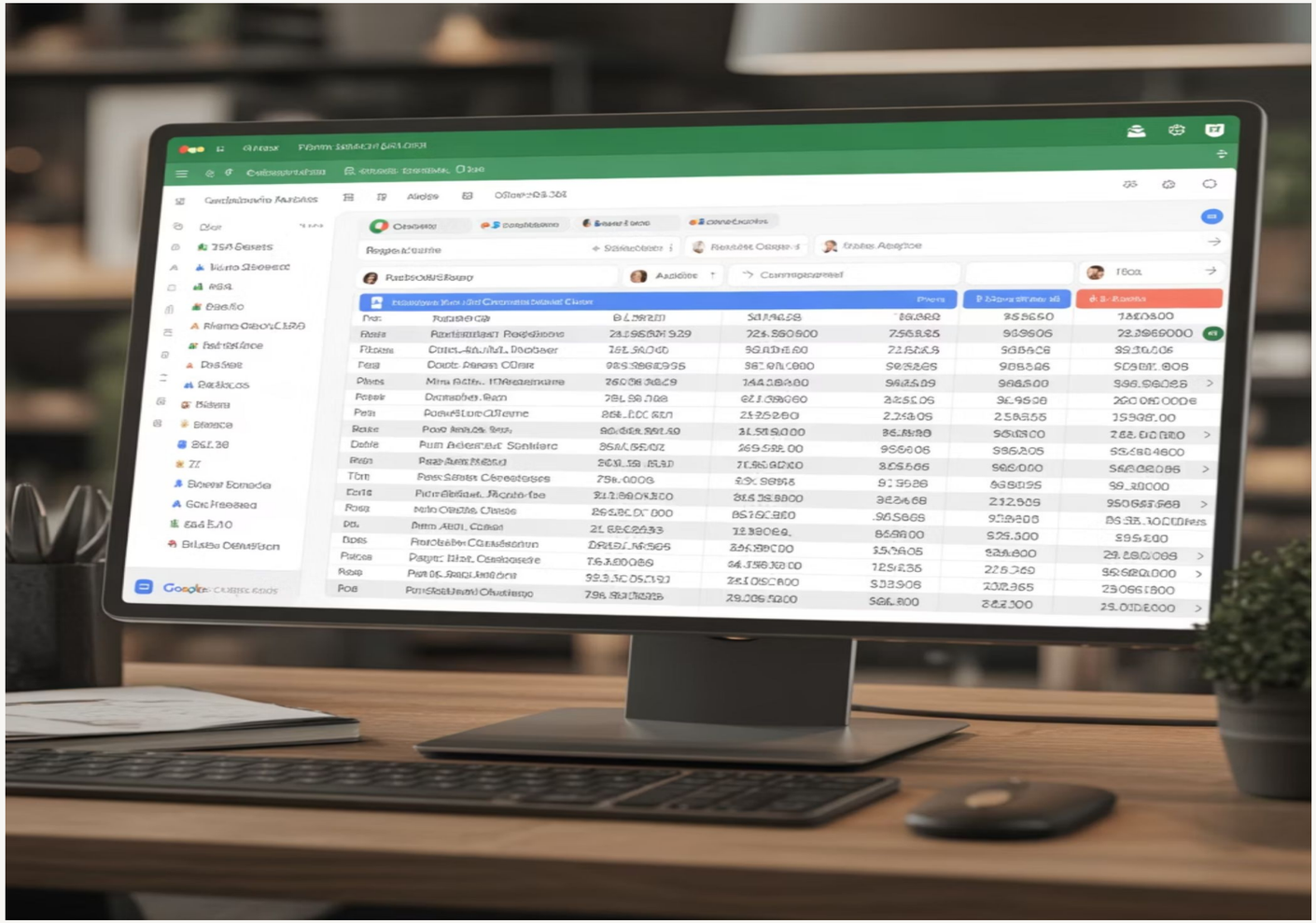
Google Sheets

Click Path:

File → Import → Upload → Replace Spreadsheet

Capture:

Screenshot of first 10–15 rows.



Powered by Futurense

Step 2: Clean & Prepare Data

Tool:

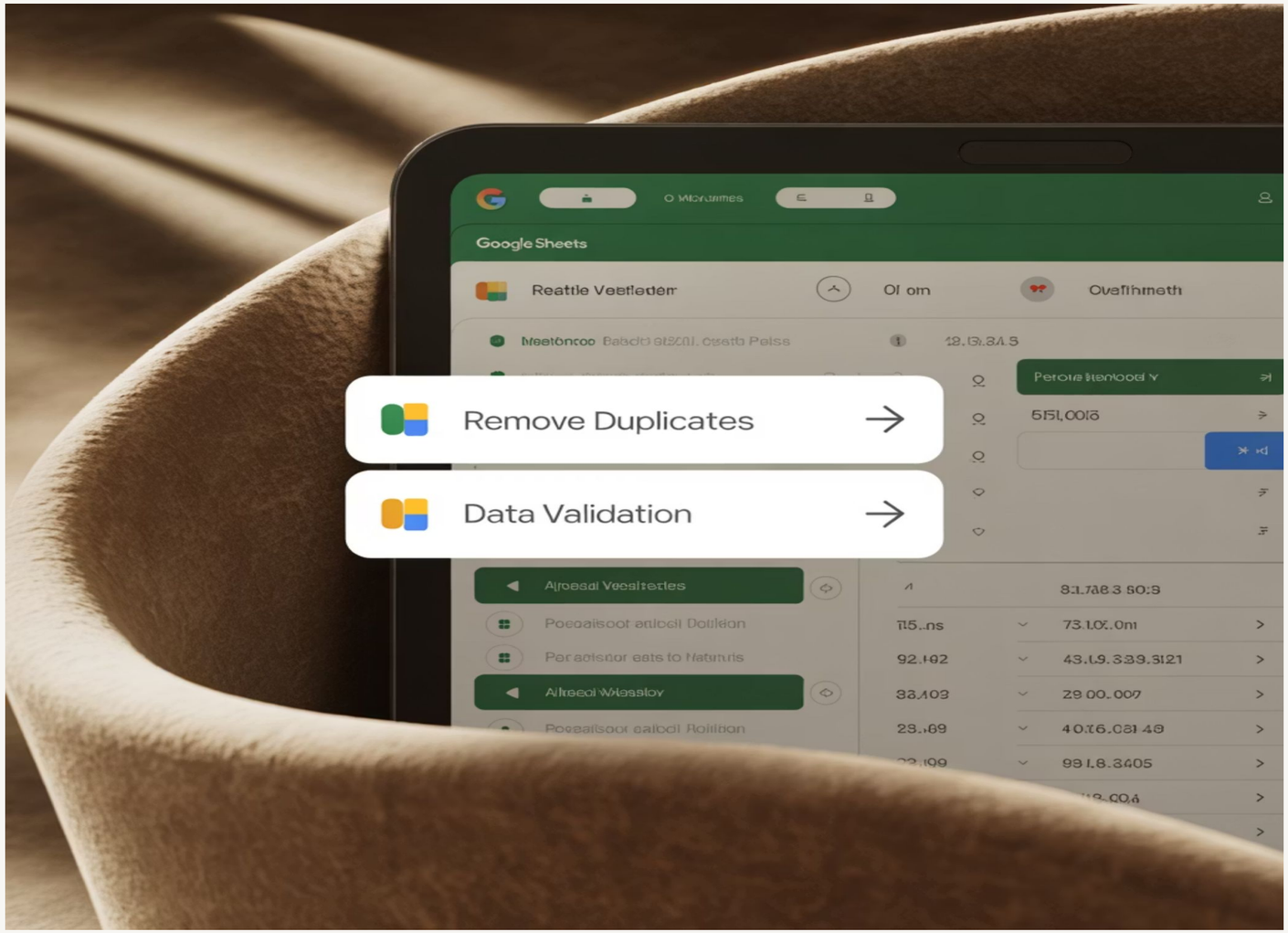
Google Sheets

Click Path:

- Data → Data Cleanup → Remove duplicates
- Format → Number
- Data → Data Validation
- File → Download → CSV

Capture:

Screenshots of data cleanup steps.



Powered by Futurense

Step 3: Convert CSV to ARFF

Tool:

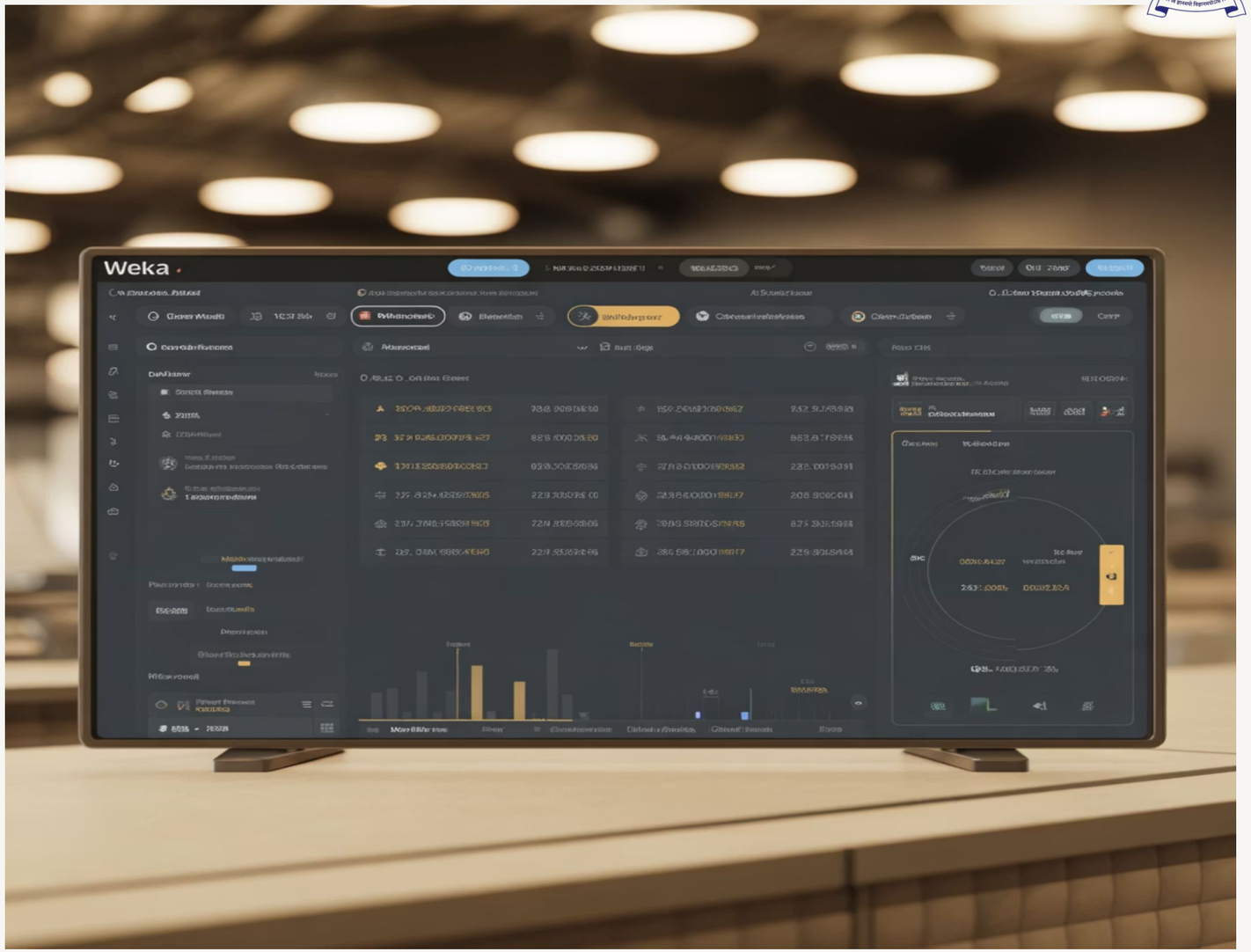
Weka → ARFF Viewer

Click Path:

Weka → Tools → ArffViewer → File → Open → Save As → .arff

Capture:

Screenshot of ARFF Viewer with attributes.



Powered by Futurense

Step 4: Load Dataset in Weka

Tool:

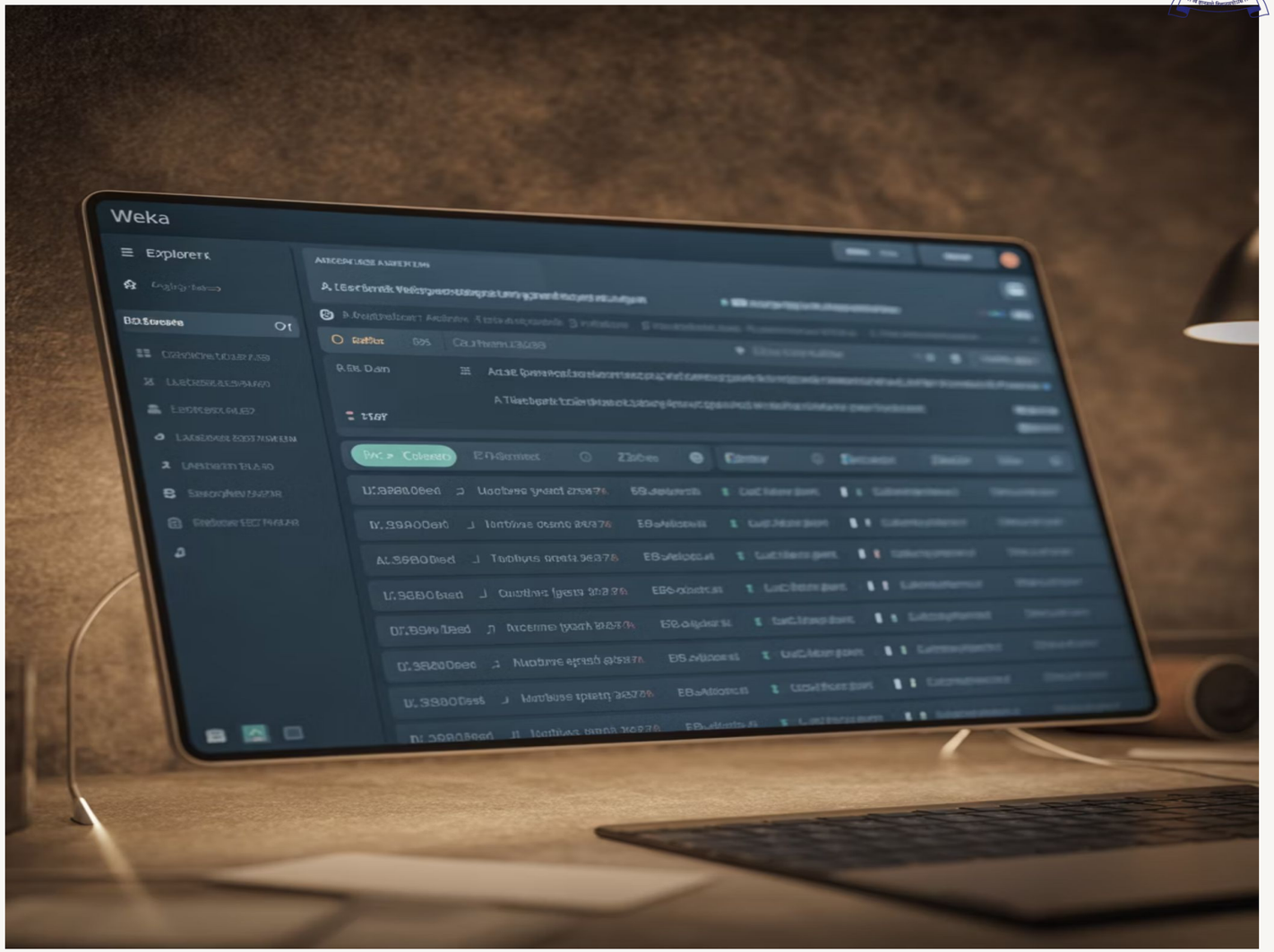
Weka → Explorer

Click Path:

Explorer → Preprocess → Open File

Capture:

Screenshot showing all attributes and class variable.



Powered by Futurense

Step 5: Explore Attributes

Tool:

Weka → Explorer

Click Path:

Preprocess → Select Attribute → Visualize All

Capture:

Screenshot of graphs:

- Attrition vs Job Satisfaction
- Attrition vs OverTime

Powered by Futurense

Step 6: Create Visual Narratives

Tool:

Flourish

Click Path:

New Visualization → Upload Data → Choose Chart

Create:

- Attrition by Department
- Attrition vs JobSatisfaction

Capture:

Final chart screenshots.

Powered by Futurense

Step 7: Train Baseline Model

Tool:

Weka

Click Path:

Classify → Choose → Trees → J48 → Use Training Set → Start

Capture:

Screenshot of:

- Accuracy
- Confusion Matrix

Overfitting Activity (1/2)

Purpose:

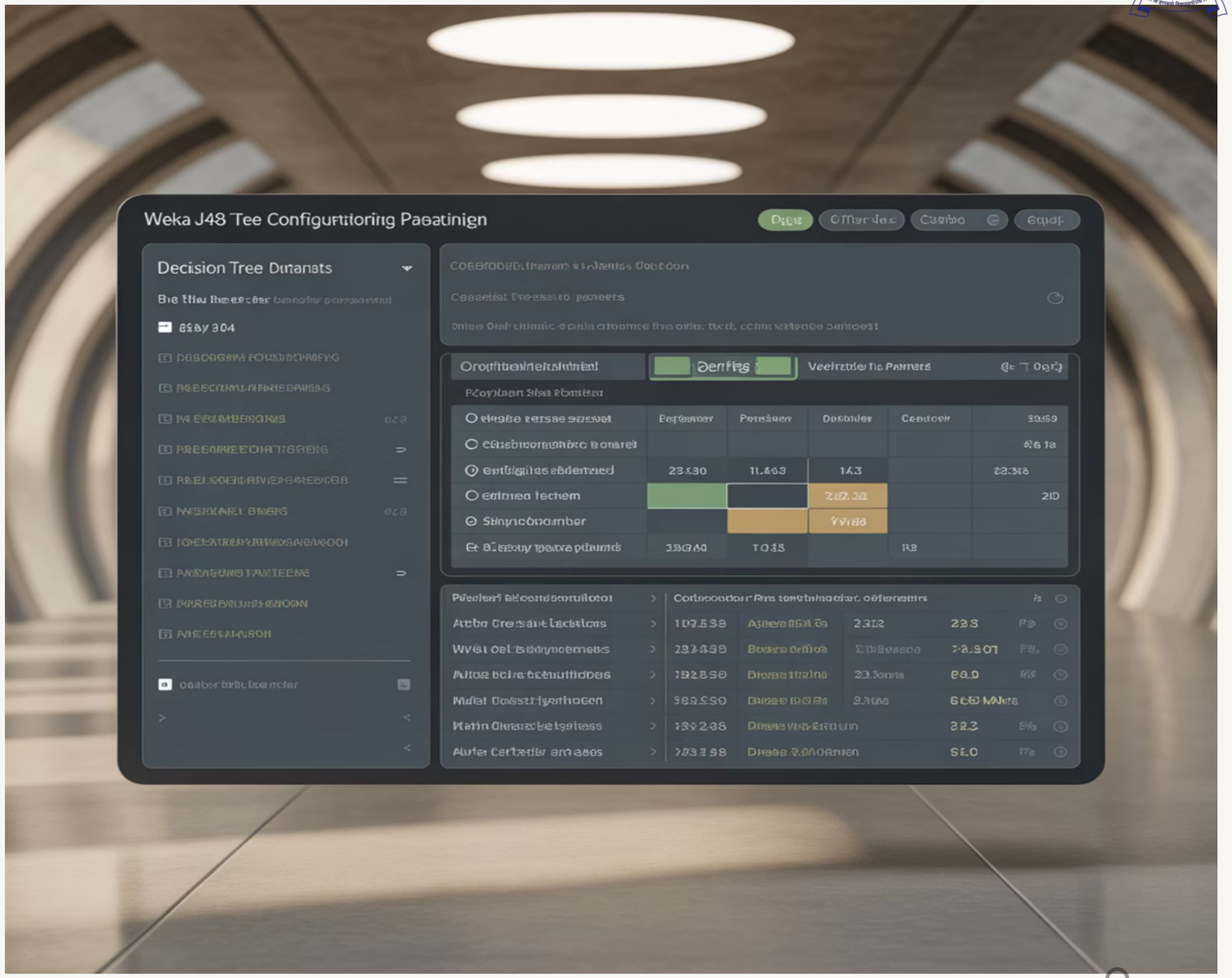
Create an overfitted decision tree.

Tool:

Weka

Click Path:

Classify → Trees → J48 → More Options → minNumObj=1 & unpruned=True → Start



Powered by Futurense

Overfitting Activity (2/2)

Observe:

- Training accuracy near 100%
- Cross-validation accuracy drops

Capture:

- Training accuracy screenshot
- Cross-validation accuracy screenshot

Powered by Futurense

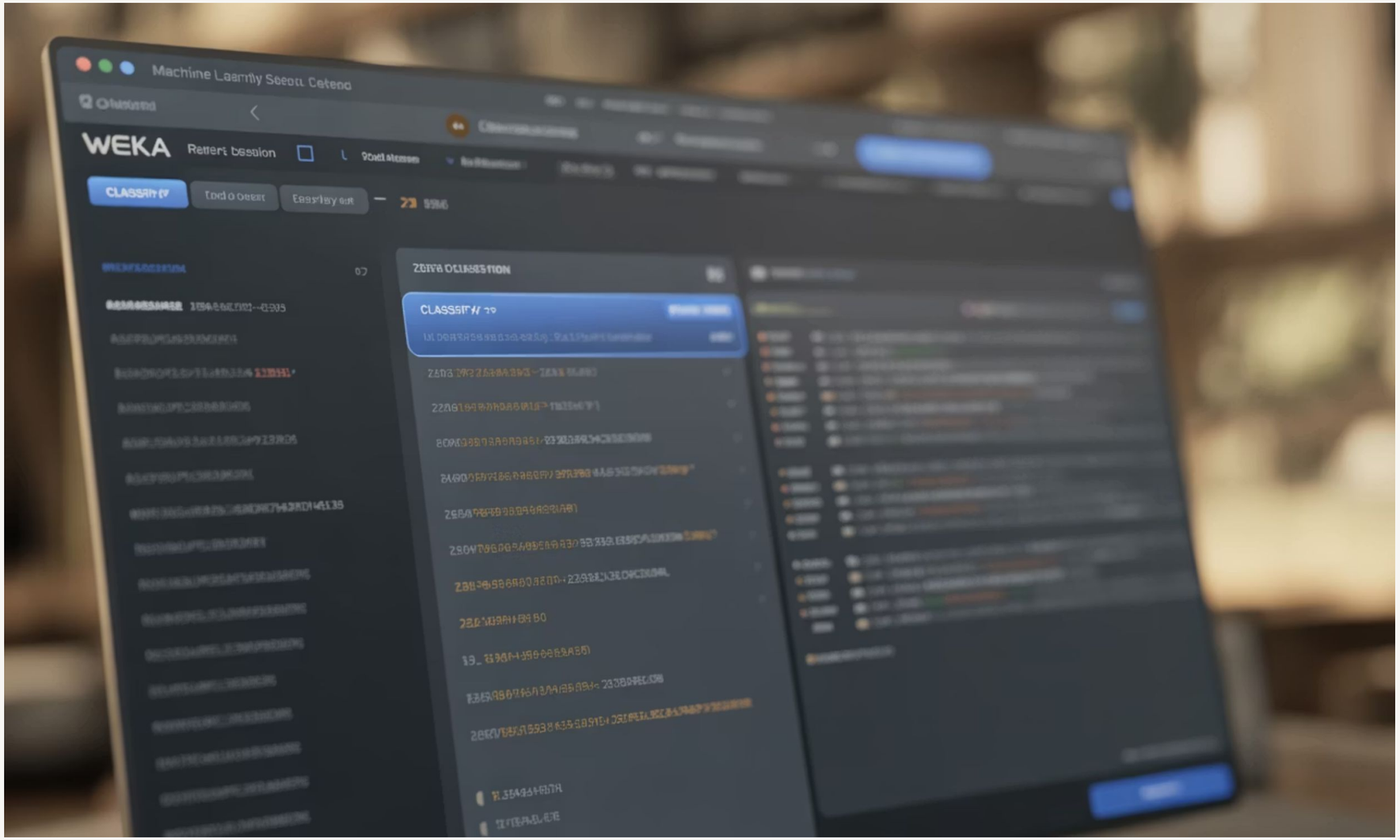
Slide 19 — Underfitting Activity (1/2)

Purpose: Build a too-simple model.

Tool: Weka

Click Path:

Classify → Rules → ZeroR → Start



Powered by Futurense

Slide 20 — Underfitting Activity (2/2)

Observe:

- Majority class prediction
- Low overall accuracy

Capture:

- ZeroR model output screenshot
- Evaluation screenshot

Powered by Futurense

Slide 21 — Step 10: Summarize Insights

Tool: Napkin AI

Paste model outputs: "Summarize insights and recommend improvements."

Capture: Summary screenshot.

Powered by Futurense

Slide 22 — Clues & Hints

Overfitting

If training accuracy >> test accuracy → Overfitting

Underfitting

If accuracy is low everywhere → Underfitting

Generalization

Simpler models usually generalize better

Look for strong predictors:

- Overtime
- Job Satisfaction
- Promotions

Powered by Futurense

Slide 23 — Learning Outcomes

Students will learn to:

01

Perform end-to-end ML workflow

02

Clean data

03

Visualize insights

04

Build classification models

05

Diagnose overfitting/underfitting

06

Interpret HR-related predictions

Powered by Futurense

Slide 24 — Submission Checklist

Students must combine all screenshots into a single PDF:

Submit:

- Google Sheets cleaning screenshots
- Flourish visualizations
- Weka ARFF Viewer
- Baseline model output
- Overfitting model output
- Underfitting model output
- Napkin AI summary
- Short written reflection

Powered by Futurense

Reflection Questions

Answer any 3:

1. Which factors contributed most to attrition?
2. Did your model overfit or underfit? Why?
3. How can dataset quality be improved?
4. How should HR teams use this model?

Powered by Futurense

Thank You

Powered by Futurense