CRISP-DM in Action

Scenario: Improving Student Retention in a University

- A university wants to identify students at risk of dropping out so it can intervene early.

1. **Business Understanding**
2. **Data Understanding**
3. **Data Preparation**
4. **Modeling**
5. **Evaluation**
6. **Deployment**

1 Business Understanding

Goal:

- Reduce student dropout rate by 15% next academic year.
- Business Questions:
- Which students are most likely to drop out?
- What factors contribute most to dropout?

Success Criteria:

- Predict at-risk students with at least 80% accuracy
- Actionable insights for academic advisors

2 Data Understanding

Data Collected:

- Student demographics (age, gender, location)
- Academic performance (GPA, failed subjects)
- Attendance records
- LMS activity (logins, submissions)
- Financial status (scholarship, unpaid fees)

Initial Analysis:

- Check missing values
- Identify outliers (very low GPA, zero attendance)
- Explore patterns (e.g., low LMS activity)

3 Data Preparation

📊 Finding: Students with low attendance and declining GPA tend to drop out.

Actions Taken:

- Remove duplicate records
- Handle missing GPA values (mean or median)
- Convert categorical data (e.g., scholarship: Yes/No → 1/0)
- Create new features:
    - Attendance rate (%)
    - GPA trend (improving or declining)

Final Dataset:

- Clean, structured data ready for modeling.

4️⃣ Modeling

Techniques Used:

- Logistic Regression
- Decision Tree
- Random Forest

Example Model Output:

- Probability a student will drop out:
    - Student A → 82% risk
    - Student B → 18% risk

📌 **Key Predictors Identified:**

- Attendance rate
- GPA trend
- LMS engagement

5️⃣ Evaluation

Model Performance:

- Accuracy: 85%
- Precision & Recall reviewed
- Confusion matrix checked

Business Check:

- Does the model identify students early enough?
- Are results understandable by staff?

✅ Result: Model meets business and technical goals.

6️⃣ Deployment

How Results Are Used:

- Dashboard for academic advisors
- Monthly list of at-risk students
- Automated alerts for intervention

Actions Taken:

- Counseling sessions
- Academic support programs
- Financial aid guidance

📈 Impact: Dropout rate reduced by 17%.

📌 CRISP-DM Summary Table

Phase	What Happens
Business Understanding	Define problem & goals
Data Understanding	Explore and assess data
Data Preparation	Clean & transform data
Modeling	Apply analytics models
Evaluation	Check accuracy & usefulness
Deployment	Use insights in real life