

Reporting Data Mining Projects in a Business Context

A Professional Reference Guide for Level 6 and Level 7 Students

Introduction

Strong data mining work can fail if it is poorly reported. In business environments, decision-makers often judge analytics projects not only by technical quality, but by how clearly findings are communicated, how well recommendations are justified, and how confidently risks are explained.

This guide is designed to help you produce professional-quality reports for data mining and business analytics projects. It is based on the CRISP-DM framework (Cross-Industry Standard Process for Data Mining), widely used in commercial analytics work.

This is **not** a marking rubric. It is a guide to good practice in communicating analytical work effectively.

Reporting using CRISP-DM

A well-structured report often follows the same logic as the CRISP-DM lifecycle:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modelling
5. Evaluation
6. Deployment / Recommendations

This gives your report a clear narrative: *What problem existed, what data was available, what was done, what was discovered, and what should happen next.*

Business Understanding

What Should Be Reported

- The business problem or opportunity
- Why the issue matters
- Stakeholders involved
- Objectives of the analysis
- Measures of success (KPIs)
- Constraints (time, budget, legal, technical)

Why It Matters

If the reader does not understand the business purpose, the rest of the report loses value. Analytics should answer a business need, not exist for its own sake.

Common Mistakes

- Starting with algorithms instead of the problem
- Vague aims (“improve performance”)
- No measurable success criteria
- Ignoring organisational constraints

Good vs Bad Example

Weak Reporting	Strong Reporting
We analysed customer data using machine learning.	We developed a churn prediction model to identify customers likely to cancel within 30 days, enabling targeted retention offers.
We wanted to improve sales.	The objective was to increase campaign conversion rates by identifying high-probability responders.

Practical Tips

- Use measurable language: reduce, increase, predict, identify
- State why the project matters commercially
- Explain who will use the findings

Data Understanding

What Should Be Reported

- Data sources
- Dataset size and structure
- Key variables
- Time period covered
- Initial observations
- Data quality issues

Why It Matters

Readers need confidence that the data is relevant, sufficient, and trustworthy.

Common Mistakes

- Listing rows/columns with no interpretation
- Ignoring missing values
- No explanation of where data came from
- No consideration of bias

Good vs Bad Example

Weak Reporting	Strong Reporting
The dataset had 10,000 rows and 15 columns.	The dataset contained 10,000 customer records covering the last 24 months, including tenure, spend, complaints, and cancellation status.
Some values were missing.	Approximately 8% of income values were missing, concentrated in older accounts, which may affect representativeness.

Practical Tips

- Explain what the variables mean
- Identify any limitations early
- Use a simple summary table

Data Preparation

What Should Be Reported

- Cleaning steps
- Handling missing values
- Removing duplicates or errors
- Transformations and encoding
- Feature engineering
- Train/test split

Why It Matters

Preparation decisions can significantly affect model quality.

Common Mistakes

- “The data was cleaned” with no detail
- Hiding preprocessing steps
- No explanation of feature choices

Good vs Bad Example

Weak Reporting	Strong Reporting
The data was cleaned and prepared.	Missing income values were imputed using median values. Categorical variables were one-hot encoded for modelling.
Variables were selected.	Ten predictor variables were retained after removing identifiers and highly correlated duplicates.

Practical Tips

- Be specific
- Explain why changes were made
- Ensure work is reproducible

Modelling

What Should Be Reported

- Algorithms tested
- Why those models were chosen
- Parameter settings
- Training process
- Comparison of candidate models

Why It Matters

Readers need to know that modelling choices were sensible and justified.

Common Mistakes

- Naming one model without explanation
- Using complex methods unnecessarily
- No comparison between alternatives

Good vs Bad Example

Weak Reporting	Strong Reporting
We used a decision tree.	Decision trees were tested due to their interpretability for managers and suitability for mixed data types.
The random forest was best.	Random forest achieved the highest recall while maintaining acceptable precision, outperforming logistic regression and decision trees.

Practical Tips

- Explain choices in business terms
- Consider interpretability, speed, maintainability
- Compare alternatives where possible

Evaluation

What Should Be Reported

- Results on unseen data
- Relevant metrics
- Comparison to baseline
- Reliability and limitations
- Suitability for use

Why It Matters

A technically accurate model may still be unsuitable in practice.

Common Mistakes

- Reporting training accuracy only
- Using the wrong metric
- No baseline comparison
- Claiming success without caveats

Good vs Bad Example

Weak Reporting	Strong Reporting
Accuracy was 91%, so the model was successful.	Test accuracy was 91%; however recall for fraud cases was only 54%, indicating many fraudulent transactions would still be missed.
The model worked well.	The model outperformed the existing rules-based system by 12 percentage points on recall.

Practical Tips

- Ask: what does success mean here?
- Use metrics linked to business risk
- Discuss limitations honestly

Deployment / Recommendations

What Should Be Reported

- How findings should be used
- Who should use them
- Operational integration
- Monitoring needs
- Risks and next steps

Why It Matters

Reports should lead to action.

Common Mistakes

- Ending at model metrics
- No recommendations
- Unrealistic implementation claims

Good vs Bad Example

Weak Reporting	Strong Reporting
The model performed well.	Recommend deploying the model as a weekly lead-prioritisation tool for the sales team.
Further work may help.	Retrain quarterly and monitor drift as customer behaviour changes seasonally.

Practical Tips

- Be realistic
- Make clear recommendations
- Link findings to decisions

Writing in a Business Style

Writing for Managers vs Technical Readers

Managers often want:

- Main findings
- Financial impact
- Risks
- Recommendations

Technical readers may also want:

- Methods
- Parameters
- Validation details

Good Practice Structure

1. Executive summary
2. Problem and objectives
3. Data and methods
4. Results
5. Recommendations
6. Appendices

Clarity and Brevity

Use:

- Short paragraphs
- Clear headings
- Direct sentences

Avoid:

- Long unexplained jargon
- Repetition
- Overly academic padding

Example

Weak	Strong
Utilisation of predictive methodologies facilitated optimisation potential.	Predictive modelling identified customers most likely to churn.

Using Tables, Charts, and Appendices

Use tables for:

- Comparing models
- Summarising variables
- Reporting metrics

Use charts for:

- Trends over time
- Distributions
- Segment comparisons

Use appendices for:

- Full confusion matrices
- Code snippets
- Extra outputs
- Detailed parameter settings

Main body = insight

Appendix = supporting detail

Reporting Metrics Properly

Classification Tasks

Use depending on context:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

Example

Fraud detection often values **recall** more than accuracy.

Prediction / Forecasting Tasks

Use:

- MAE (Mean Absolute Error)
- RMSE (Root Mean Square Error)
- MAPE (if appropriate)

Business Context Matters

Metric	Business Meaning
Recall	How many true cases were found
Precision	How many alerts were worthwhile
RMSE	Typical forecast error size

Never report metrics without interpretation.

Ten Signs of a Weak Data Mining Report

1. No clear business objective
2. Starts with tools instead of the problem
3. Poor explanation of data sources
4. No discussion of missing data or bias
5. Vague preprocessing description
6. Model chosen without justification
7. Accuracy reported with no context
8. No limitations acknowledged
9. No recommendations given
10. Difficult to read

Ten Habits of Strong Analytics Reporting

1. Clear purpose from page one
2. Business language as well as technical language
3. Transparent data description
4. Honest discussion of quality issues
5. Logical structure
6. Justified modelling choices
7. Appropriate metrics
8. Realistic interpretation
9. Actionable recommendations
10. Professional presentation

Example Paired Excerpts

Example A — Opening Paragraph

Weak

We used Weka to run machine learning algorithms on the dataset.

Strong

This project aimed to reduce customer churn by identifying accounts likely to cancel within 30 days. Predictive models were developed and compared using historical customer behaviour data.

Example B — Results Paragraph

Weak

The random forest was best with 89%.

Strong

The random forest achieved 89% test accuracy and substantially improved recall of at-risk customers compared with logistic regression, making it the strongest candidate for retention targeting.

Example C — Recommendation Paragraph

Weak

The company should use the model.

Strong

It is recommended that the model be deployed as a weekly prioritisation tool for retention staff, with human review of flagged high-risk customers.

Final Checklist For Data Mining Assignments

Before you submit your report, ask:

Business Understanding

- ☐ Is the problem clear and specific?
- ☐ Have I explained why it matters?

Data Understanding

- ☐ Are data sources explained?
- ☐ Have I discussed quality issues?

Preparation

- ☐ Have I described cleaning and transformation steps clearly?

Modelling

- ☐ Have I justified my modelling choices?

Evaluation

- ☐ Have I used suitable metrics?
- ☐ Have I interpreted results in business terms?
- ☐ Have I acknowledged limitations?

Recommendations

- ☐ Have I stated what should happen next?
- ☐ Are recommendations realistic?

Communication

- ☐ Is the report easy to navigate?
- ☐ Are tables and charts clear?
- ☐ Is the writing concise and professional?

Final Thought

A strong analytics report does more than describe what was done. It demonstrates judgement, builds trust, and helps an organisation decide what to do next.