



Credit Risk Assessment for GE Using CRISP-DM

Welcome to our comprehensive guide on implementing a data mining project using the CRISP-DM methodology. This presentation will walk you through each phase, from business understanding to deployment.

By Solar Shrestha



CRISP-DM Overview



- 1 Business Understanding
Define objectives and problem statement
- 2 Data Understanding
Explore and verify data
- 3 Data Preparation
Create analytic architecture
- 4 Modeling
Apply data mining techniques
- 5 Evaluation
Assess model performance
- 6 Deployment
Implement solution in business context



Image Source: <https://learn.snhu.edu/d2l/le/content/1739062/viewContent/36029637/View>



Business Understanding Phase



Problem Statement

Clearly define the business problem to be solved.



Business Objectives

Identify specific, measurable goals for the project.



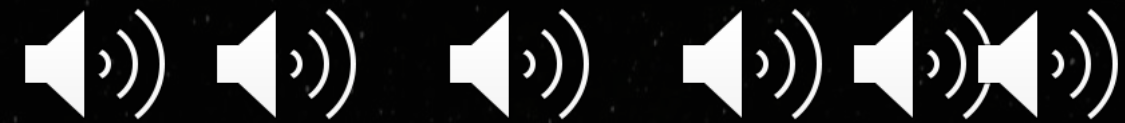
Industry Context

Understand the organization's position and challenges in the industry.



Data Mining Goals

Translate business objectives into data mining goals.



Data Understanding Phase

Describe Data

Identify data sources and types.
Document data characteristics and formats.

Explore Data

Perform initial data analysis. Use descriptive statistics to understand data distribution.

Verify Data

Check data quality and completeness.
Identify potential issues or biases.

Data Preparation: Analytic Architecture

1

Data Collection

Gather data from various sources

2

Data Integration

Combine data into a unified format

3

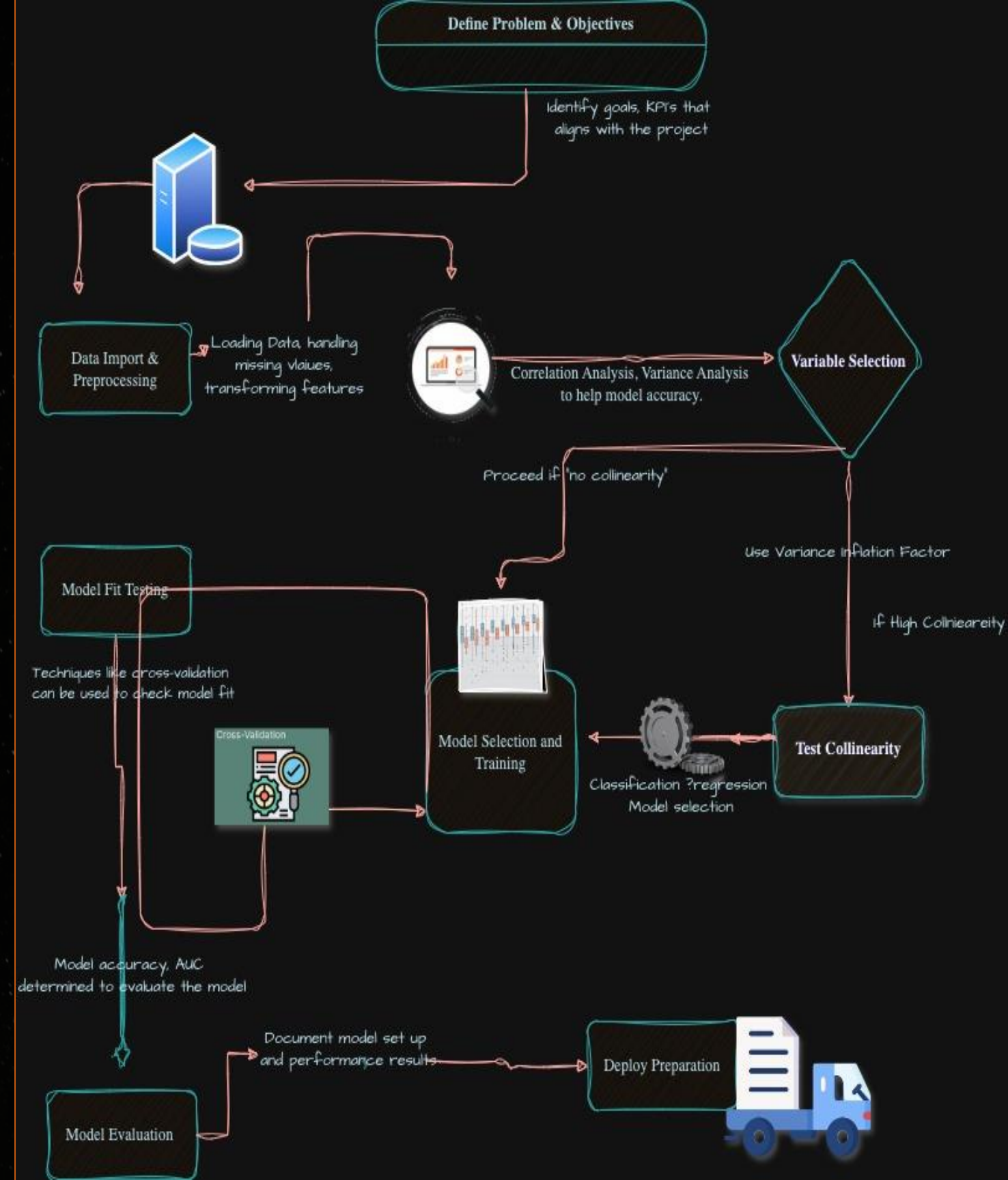
Data Cleaning

Address missing values and outliers

4

Feature Engineering

Create new variables to improve model performance



Data Quality and Integrity

Accuracy

Ensure data correctness and precision

Completeness

Address missing or incomplete data points

Consistency

Maintain uniform data formats across sources

Timeliness

Use up-to-date and relevant data





Data Protection Strategies



Encryption

Protect sensitive data during storage and transmission



Access Control

Implement role-based access to data resources



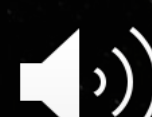
Anonymization

Remove personally identifiable information from datasets

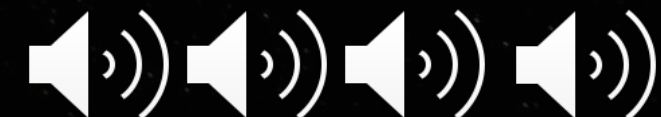
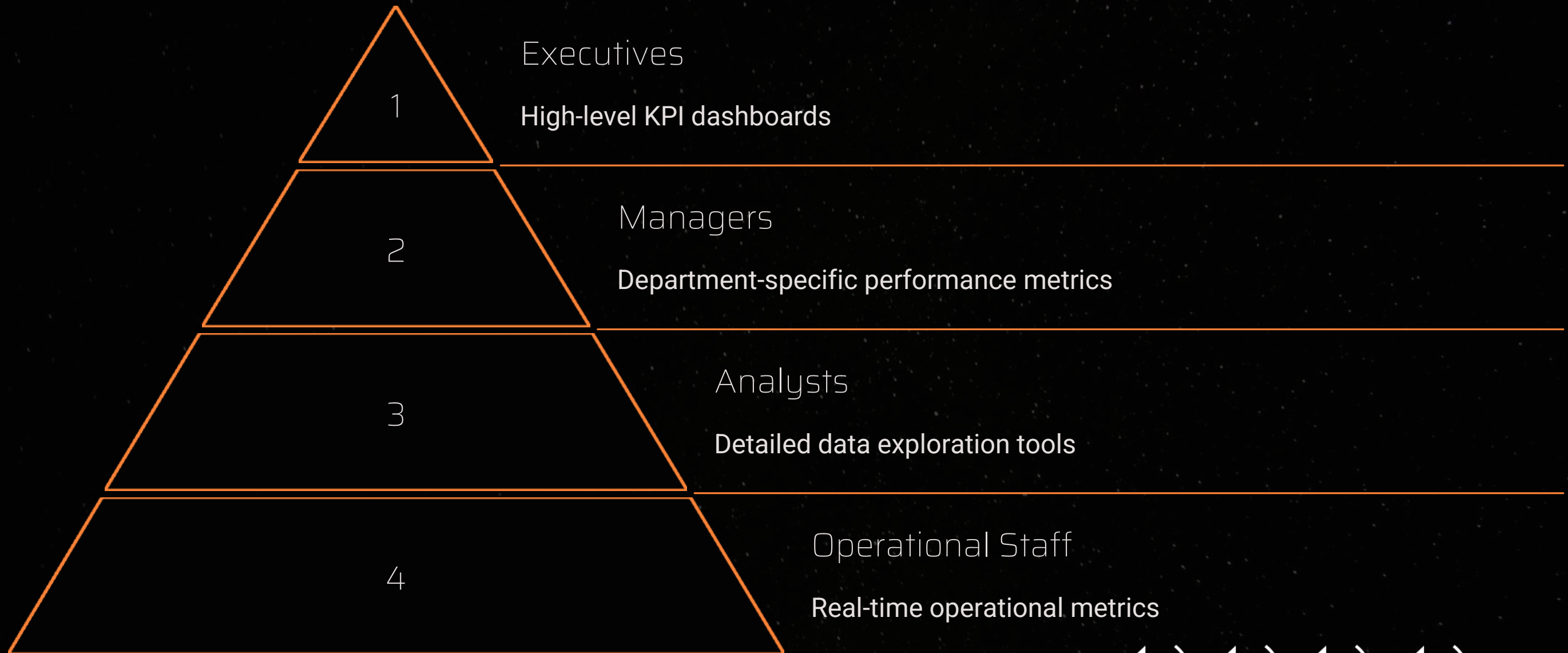


Data Retention

Establish policies for data storage and deletion



Stakeholder Visualization Needs



Leveraging Industry Knowledge for Data-Driven Success



Harness deep industry insights to identify untapped data sources. Transform raw information into actionable intelligence for your organization.

Enhance Competitiveness

Gain a strategic edge by leveraging unique industry data.

Boost Effectiveness

Streamline operations and decision-making with data-driven insights.

Ensure Longevity

Future-proof your organization through adaptive data strategies.



Deployment Plan

1

Solution Testing

Validate model performance in controlled environment

2

Pilot Implementation

Deploy solution to limited user group

3

Full Rollout

Expand solution to entire organization

4

User Training

Provide comprehensive training to end-users

Project Impact and Conclusions

25%

Efficiency Increase

Streamlined processes led to significant time savings

95%

User Satisfaction

High adoption rate and positive feedback from stakeholders

\$1M

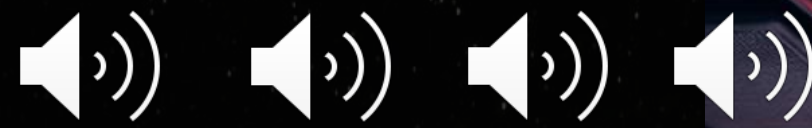
Cost Savings

Reduced operational costs through data-driven decision making

3x

ROI

Project delivered substantial return on investment



Questions and Feedback Session



Reference

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, 41(3), 1-52. <https://doi.org/10.1145/1459352.1459353>

Chen, D., & Zhao, H. (2012). Data security and privacy protection issues in cloud computing. *International Conference on Computer Science and Electronics Engineering*, 647-651. <https://doi.org/10.1109/ICCSEE.2012.193>

NIST. (2020). Guide to protecting the confidentiality of personally identifiable information (PII): Special Publication 800-122. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-122>

Olson, J. E. (2003). *Data quality: The accuracy dimension*. Morgan Kaufmann.

Stallings, W. (2017). *Cryptography and network security: Principles and practice (7th ed.)*. Pearson.

Tankard, C. (2012). Big data security. *Network Security*, 2012(7), 5-8. [https://doi.org/10.1016/S1353-4858\(12\)70063-6](https://doi.org/10.1016/S1353-4858(12)70063-6)

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 623-640. <https://doi.org/10.1109/69.404034>

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a service and big data. *International Conference on Advances in Cloud Computing (ACC)*, 21-29.

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring (2nd ed.)*. Analytics Press.

Tableau Software. (2020). Visual analysis best practices: Simple techniques for making every dashboard more impactful. Tableau Software. Retrieved from <https://www.tableau.com>

Marr, B. (2018). Data-driven decision making: 10 simple steps for business success. *Forbes*. Retrieved from <https://www.forbes.com/sites/bernardmarr/2018/01/29/data-driven-decision-making-10-simple-steps-for-business-success/>

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business Press.

Wang, J., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5-33. <https://doi.org/10.1080/07421222.1996.11518099>

Stallings, W. (2017). *Cryptography and network security: Principles and practice (7th ed.)*. Pearson.

NIST. (2020). Guide to protecting the confidentiality of personally identifiable information (PII): Special Publication 800-122. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-122>

Chen, D., & Zhao, H. (2012). Data security and privacy protection issues in cloud computing. *International Conference on Computer Science and Electronics Engineering*, 647-651. <https://doi.org/10.1109/ICCSEE.2012.193>

Tankard, C. (2012). Big data security. *Network Security*, 2012(7), 5-8. [https://doi.org/10.1016/S1353-4858\(12\)70063-6](https://doi.org/10.1016/S1353-4858(12)70063-6)

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a service and big data. *International Conference on Advances in Cloud Computing (ACC)*, 21-29.

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring (2nd ed.)*. Analytics Press.