MERGING LANES: WHERE E-LEARNING DIVERSITY MEETS FUTURE TRENDS

VOLUME 11, 2026
e-ISBN : 978-629-98755-9-8

eISBN 978-629-98755-9-8
9 786299 875598

SIG CS@e-Learning
Unit Penerbitan
Jabatan Sains Komputer & Matematik
Kolej Pengajian Pengkomputeran, Informatik & Matematik
Universiti Teknologi MARA Cawangan Pulau Pinang

MERGING LANES: WHERE E-LEARNING DIVERSITY MEETS FUTURE TRENDS

Advisor
Dr. Nor Hanim Abd Rahman,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

Chief Editor
Ts. Jamal Othman,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

Editors
Ts. Dr. Syarifah Adilah Mohamed Yusoff,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

Dr Arifah Fasha Rosmani,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

Mohd Saifulnizam Abu Bakar,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

Published by:
Unit Penerbitan Jabatan Sains Komputer & Matematik (JSKM)
Bahagian Hal Ehwal Akademik (BHEA)
Universiti Teknologi MARA
Cawangan Pulau Pinang
13500 Permatang Pauh
Pulau Pinang
Malaysia

e ISBN : 978-629-98755-9-8

DATA ANALYTIC FOR BUSINESS BASED PREDICTION

*Syarifah Adilah Mohamed Yusoff[1], Elly Johana Johan[2], Saiful Nizam Warris[3] and Jamal Othman[4]
*syarifah.adilah@uitm.edu.my[1], ellyjohana@uitm.edu.my[2], saifulwar@uitm.edu.my[3], jamalothman@usm.edu.my[4]

[1,2,3,4]Jabatan Sains Komputer & Matematik (JSKM),
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

*Corresponding author

ABSTRACT

Currently, business processes are perceived not merely as a sequence of activities responding to an event to generate output, but as a complex system involving the interplay of individuals, technologies, strategies, and business rules to attain certain business outcomes. Consequently, the analysis of a substantial volume of data is essential not only for present operations and several years ahead but also for future trends and long-term objectives. This study aims to present the concept of data analytics within the business domain, integrating it with a business framework specifically for operational purposes, and incorporating machine learning for predictive analytics, culminating in the evaluation of classification predictions. Information is a crucial asset that enables future business planning through a data-driven methodology and demonstrates the importance of business analytics for future success.

Keywords: Business Process, Data Analytics, Machine Learning, Predictive Analytics, Business Intelligence

Introduction

In the era of digital transformation, organizations generate vast amounts of data from operational systems, online transactions, and customer interactions. Transforming this raw data into actionable insights has become a strategic priority for organizations seeking competitive advantage. As a result, Business Intelligence has emerged as an important approach for collecting, integrating, and visualizing organizational data to support managerial decision making (Abusweilem & Abualous, 2019; Delen & Zolbanin, 2018).

While Business Intelligence primarily focuses on descriptive analysis and historical reporting, organizations increasingly rely on Business Analytics to perform deeper analytical processes that support strategic and operational decisions. Business analytics incorporates statistical analysis, data modelling, and computational techniques to extract insights from complex datasets and support data-driven decision making (Suri, 2021; Tsai et al., 2015).

A critical component of business analytics is Data Mining, which focuses on discovering hidden patterns, relationships, and knowledge from large volumes of data. Data mining techniques such as classification, clustering, and association rule learning allow organizations to uncover meaningful insights that may not be visible through traditional analytical approaches (Liao et al., 2012; Hall et al., 2022; Fayyad et al., 1996).

To ensure that data mining projects are conducted systematically and effectively, structured methodologies are often adopted. One widely used methodology is the Cross-Industry Standard Process for Data Mining (CRISP-DM), which provides a comprehensive framework consisting of business understanding, data understanding, data preparation, modelling, evaluation, and deployment phases (Chapman P., 2000).

Building upon these analytical foundations, Predictive Analytics enables organizations to forecast future trends, customer behaviour, and operational outcomes based on historical data. Predictive analytics commonly utilizes machine learning algorithms and statistical models to identify patterns that can be used to predict future events and support proactive decision making (Shmueli & Koppius, 2011; Siegel, 2020).

Given the increasing importance of predictive capabilities in modern organizations, this study explores the implementation of business analytics using the CRISP-DM framework to support predictive modelling. The research further examines the application of machine learning algorithms, analytical tools, and visualization techniques in modelling and analysing predictive outcomes for business decision support.

CRISP-DM as a framework for business analytic

CRISP-DM defined the framework that data mining professionals first use to plan their data mining projects. In 1996, it was first revealed by the NCR Corporation and OHRA for the first time to provide the data and mining community with their first business-oriented, repeatable, and reliable framework (Wirth & Hipp, 2000). The CRISP-DM is considered as the most structured, iterative, and domain-neutral methodology to derive and implement a framework for a data-driven project. The application of CRISP-DM is still widely accepted and followed in the fields of data mining and predictive analytics. The CRISP-DM was built with the flexibility to span across multiple projects and therefore was adapted into the education and financial assistance systems, among many other industries. In CRISP-DM, there are 6 phases that span across business and data understanding, data preparation, model building, and evaluation and deployment, which are the most common data mining business processes to present in a process framework.

Figure 1 shows that business understanding starts with project goals definition and aligning them with the organizational priorities for the project to succeed. Further Table 1 elaborates starting from top level of understand the business and next drill out the data in data understanding phase, where the raw data is collected and explored to judge data quality and find patterns or anomalies. The most labour intensive phase of data science is data preparation, where data is cleaned, transformed and structured to make a usable dataset. After the data is ready, during the modelling usually statistical and machine learning techniques used to refine models so that they best perform. In the evaluation phase, the developed models are compared to the predefined criteria to ensure that they deliver against the business objectives. Finally, the results are deployed into decision making processes like generating reports, or automating the workflows, so that it can be put into practice and continuously improved for future projects (Schröer et al., 2021).

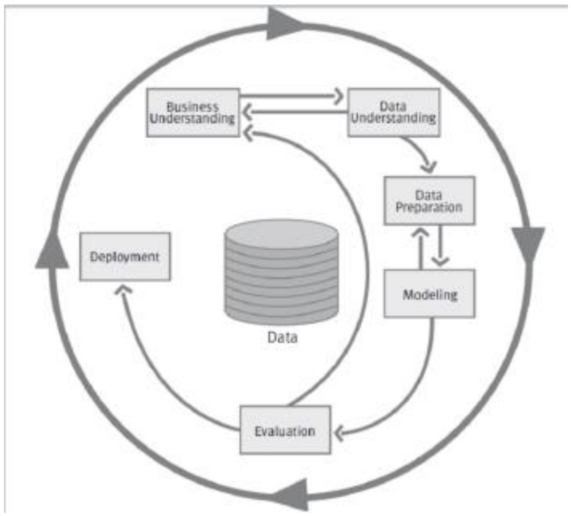

Figure 1: Phases of the CRISP-DM reference model (Huber et al., 2019; Chapman, 2000)

Table 1: The detail roles of each phase in CRISP-DM

CRISP-DM Phases	Description
1. Business Understanding	During this stage, it is important to comprehend the project's requirements and business objectives. It comprises defining the issue and determining the objectives. Gaining a thorough grasp of the business environment in which the data mining project will be implemented is the main goal.
2. Data Understanding	The origins of the data are being determined at this stage. Additionally, the initial data collecting has been completed. Consequently, it also entails looking into and comprehending the connections among the facts.
3. Data Preparation	These phases are involving several steps which is cleaning, transforming and the pre-processing the data to make it suitable for do the analysis.
4. Modelling	Various modeling techniques may be chosen and used to the prepared data during this phase. Apart from that, the project's particular modeling technique will be chosen.
5. Evaluation	In this phase, the chosen model will be assessed according to predetermined standards. Depending on the evaluation's findings and the process review, the following actions must be decided.

6. Deployment	The completed model is delivered to a production environment during the deployment phase. A final report will be generated from the deployment.

Machine learning algorithms for predictive implementation

One of the earliest components of predictive analytics involved the use of classification algorithms. These algorithms understood a priori the technique needed to predict what the different classifications of the data would be and how to organise the data into various groups, based on what patterns were learned during the training phase.

There are machine learning algorithms that are robust mechanisms that allow systems to learn from data and autonomously predict or choose without direct programming. These algorithms are designed to uncover hidden patterns and relationships in the data thereby making them important to solving difficult problems in industries such as healthcare, finance, food and beverage. Businesses can use machine learning to mine important insights, optimize processes, and enhance customer experiences - increasingly crucial to decision making in dynamic markets (Jordan & Mitchell, 2015). With additional data, machine learning is a major benefit that allows it to improve. Machine learning uses a repetitive learning approach to improve prediction accuracy and help in decision making in data driven way. They provide a solid foundation to move from intuition based to analytics-based decision making in what has to do with anticipating customer preferences, better resource allocation or predicting sales trends. Machine learning is used by businesses to deal with problems related to data accuracy and quality, optimize processes, and increase overall efficiency (Surur et al., 2025).

Classification algorithms are collection of algorithms used to predict the data based on an attribute that was known as target attribute. Random Forest, Decision Trees, and Naive Bayes are the most used classification models in varies predictive domain of studies such as healthcare, education, entertainment and many more. Naive Bayes is especially known for its effectiveness in using it for processes where the initial input was categorical (Karthikeyan & Rani, 2022). When the concept of optimal search of this algorithm is based on statistical probabilities of the event occurred among target attribute and input attributes, this algorithm has proven to be efficient for moderate range of data and less complex attributes.

Meanwhile the tree-based models most used were ID3, C4.5, and CART. They are easy to use for parameter tuning and visually interpretable (Han et al., 2011). The produced model has been used to make sense of the logic behind each prediction to validate the prediction of classification problems. The logic was necessary for validating such as an event of several classes. The models used human-like reasoning, as the trees split the data into branches using values of the attributes.

Random Forest was the first and most recognized ensemble model to build multiple trees and aggregate the prediction of each tree (Breiman, 2001). It improved the predictive accuracy and reduced overfitting. It was perfectly applicable to the data with different kinds of profiles of the applicants and highly correlated variables. The Random Forest model was shown to be one of the most applicable and dependable models for problems of financial aid classification and, applied to, realistic data.

Comparison of tools and visualisation aids for business analytic

Data mining is also crucial in elucidating the patterns of business application data as it helped to construct classifications and data visualizations. These technologies help in the construction of predictive models for business classification and in the provision of insights to assist the manager in decision-making. Different tools had varying capabilities in terms of data processing, model construction, and presentation. These tools were considered optimal for the given purpose.

Power BI is a business intelligence and data presentation tool developed by Microsoft. It is primarily used for dashboards and data visualizations. It did not directly train machine-learning models but was a great tool for the presentation of machine learning model outputs. In this case, Power BI is used to create a user-friendly interactive dashboard to represent the data and to classify and summarize apps. It has a great multi-level filtering option and interactive dashboard that to present data in the holistic output and reveal the hidden pattern for visualisation.

RapidMiner offers a visual interface with data mining capabilities and a variety of machine learning tools including Naïve Bayes, Decision Trees, and Random Forest. Users could design workflows for data prep and model training with relative ease. This made it handy for people who didn't know how to code. Because of this, it is particularly useful for developing and testing classification models in a fast and flexible way in educational settings (Han et al., 2011).

WEKA (Waikato Environment for Knowledge Analysis) is one of the first tools offered for free to the public for data mining. The University of Waikato developed it and the data mining tool became quite popular. The tool is appropriate for educational settings and small data collections as it offers a variety of algorithms, preprocessing methods, evaluation tools, and even some visualization methods (Hall et al., 2022).

Like Power BI, Tableau is also used primarily for data visualization. Tableau is popular for its capacity to make interactive charts and dashboards. Tableau also has some predictive capabilities including trend lines and forecasting, but for more advanced modelling, Tableau is used in conjunction with other tools such as R or Python. The comparision of all the above mentioned algorithms was further compared in table 2.

Table 2: Comparison table of Various Types of Data Mining Tools

Tool	Functionality	Strengths	Limitations	Suitability for
Power BI	Data visualization, dashboard reporting	- User-friendly interface - Real time interactivity - Strong integration with Excel	• Limited built-in modelling • Requires external models for prediction	Best for dashboard development and presenting model results to stakeholders
Rapid Miner	Full data mining process: data preparation, modeling, evaluation	• Drag-and-drop interface • Supports many ML algorithms • No coding needed	• Memory-intensive • May have limitations for large datasets	Suitable for building and testing predictive classification models
WEKA	Machine learning experimentation and educational tool	- Lightweight - Great for teaching and experimentation - Many built-in classification methods	• Basic visualization • Limited scalability	Ideal for models Comparison in early testing phase
Tableau	Advanced data visualization and analytics	• Visually rich dashboard designs • Strong storytelling features	• Limited modelling support • Requires external tool for advanced predictions	Alternative to Power BI for presentation; less integration with academic systems

Business intelligence systems provided users with visualisation tools that enabled them to comprehend intricate information. To facilitate the effective, timely and transparent functioning of the business case studies, visualisation tools assisted in converting the results of predictive model classifications into dashboards and reports.

Power BI, a product of Microsoft and one of the primary tools used in this initiative, it's able to combine numerous data systems and user-responsive dashboards and interactive reports containing critical charts and maps. Stedman (2020) asserts that organizations Power BI's visualization capabilities for real-time data and dynamic reporting to improve decision-making and streamline actionable insights.

Another important visualisation tool is Tableau, which has similar features but centred more on visual story creation. Tableau is appreciated more for its pleasing visual dashboard designs as well as its interactive filter designs. Tableau was used mostly in academic and corporate settings to share analyses in an engaging way. Both tools had intuitive interfaces designed to cater to the non-technical user, such as administrators, officers, and other stakeholders. More importantly, they facilitated the visibility and explainability of the data in ways that increased the efficiency and equitable allocation of business in a fair and evidence-based manner.

Modelling and Analysis of the predictive results

The modelling process could not succeed without the proper test framework, which outlined how the predictive models would be trained, evaluated, and in what ways the models' accuracy and fairness would be verified. In Table 3 the evaluation framework consists of the following components:

Table 3: Test Structure Element with Description

Element	Description
Train-Test Split	The dataset is divided into 80% and 70% for training and 20% and 30 % for testing. The training set will be used to build a model; however, the test set will be used afterwards to test the generalization capability of the model.
Cross-Validation	10-fold cross-validation is planned to improve model robustness by rotating training and validation across 10 data subsets and averaging the results.
Stratified Sampling	To address any class imbalance, stratified sampling may be used to maintain proportional representation of each zakat category in both training and test datasets.
Confusion Matrix	A matrix is used to visualize true positives, false positives, true negatives, and false negatives to assess prediction performance per zakat category.
Evaluation Metrics	Model performance is measured using classification metrics such as accuracy, precision, recall, and F1- score to evaluate predictions across categories.

In assessing the performance of the models, two primary objects are used: confusion metrics and model evaluation metrics. The confusion matrix captured the essence of the predictions of the model along the lines of true predictions and false predictions (positives/negatives) for the respective target attribute. The model provides a good basis for comparison of the actual classifications of predictors against the predictions of the model for classifications.

Furthermore, computations of metrics, for example, evaluation of total model performance, are done by model class metrics to gauge the performance of the models per class, along computation of measurement of precision, measurement of the true positives (actual true positions and false positions), and measurement of false negatives (actual true negatives and false negatives), and the model evaluation metric on the total model performance: the model evaluation metric on the false negatives (actual true negatives and false negatives/ true negatives). Table 4 show Standard Classification Evaluation Metrics for two classes of target attribute.

Table 4: Standard Classification Evaluation Metrics

Metric	Description
Accuracy	Measures the overall correctness of the model across all target attribute values (nominal) known as classes.
Precision	Indicates how many students predicted to belong to a specific class.
Recall (Sensitivity)	Shows how many actual values of specific class are correctly identified by the model.
F1-Score	Provides a balance between precision and recall; especially useful for imbalanced datasets.

Meanwhile, each metric is calculated using the following formulas in Figures 5. These statistical formulas ensured a comprehensive assessment of the model's effectiveness across all classes in target attribute in the classifications of testing and training. In case of imbalance data are expected between classes, metrics like the F1-score and class-wise precision and recall are especially important in evaluating model fairness and effectiveness in supporting classification decisions.

Figure 5: Formula for Calculating Classification Task

- **Precision for class i:**

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$$

- **Recall for class i:**

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

- **F1-Score for class i:**

$$\text{F1}_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

- **Overall Accuracy:**

$$\text{Accuracy} = \frac{\text{Total Correct Predictions}}{\text{Total Number of Predictions}}$$

Conclusion

In conclusion, the integration of Business Intelligence, Business Analytics, Data Mining, and Predictive Analytics has transformed the way organizations utilize data for strategic decision making. While Business Intelligence provides visibility into historical and operational data, advanced analytical techniques enable organizations to uncover patterns and predict future outcomes. As emerging technologies such as artificial intelligence, big data, and cloud computing continue to evolve, the role of data-driven analytics will become even more critical. Organizations that effectively adopt these analytical approaches will be better positioned to identify opportunities, mitigate risks, and maintain sustainable competitive advantages in an increasingly data-oriented business environment.

References:

Abusweilem, M. A., & Abualous, S. (2019). The impact of knowledge management process and business intelligence on organizational performance. *Management Science Letters*, 2143–2156. https://doi.org/10.5267/j.msl.2019.6.020

Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. *Journal of Business Research*, *90*, 186–195. https://doi.org/10.1016/j.jbusres.2018.05.013

Suri, N. (2021). Book review: Thomas H. Davenport and Jeanne G. Harris, Competing on Analytics. *Vision: The Journal of Business Perspective*, *25*(3), 388–389. https://doi.org/10.1177/09722629211023024

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0030-3

Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, *39*(12), 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2022). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10. https://doi.org/10.1145/1656274.1656278

Chapman, P. (2000). CRISP-DM 1.0: Step-by-step data mining guide.

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572. https://doi.org/10.2307/23042796

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–37. https://doi.org/10.1609/aimag.v17i3.1230

Siegel, E. (2020). Predictive analytics. The Business & Management Collection, 2020(5), e1008168. https://doi.org/10.69645/wfng5530

Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a Standard Process Model for Data Mining*. http://cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*(1), 526–534.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*(1), 403–408.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Surur, F. M., Mamo, A. A., Gebresilassie, B. G., Mekonen, K. A., Golda, A., Behera, R. K., & Kumar, K. (2025). Unlocking the power of machine learning in big data: a scoping survey. *Data Science and Management*. https://doi.org/10.1016/j.dsm.2025.02.004

Karthikeyan, N., & Rani, M. S. (2022). ECG Classification Using Machine Learning Classifiers with Optimal Feature Selection Methods. *Lecture Notes on Data Engineering and Communications Technologies*, 277–289. https://doi.org/10.1007/978-981-16-9605-3_19

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Han, Jiawei, Kamber, Micheline, & Pei, Jian. (2011). *Data Mining Third Edition*. https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

Stedman, C. (2020, September). *What is Business Intelligence? BI Defined*. SearchBusinessAnalytics. https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI

eISBN 978-629-98755-9-8

9 786299 875598

SIG CS@e-Learning
Unit Penerbitan
Jabatan Sains Komputer & Matematik
Universiti Teknologi MARA Cawangan Pulau Pinang

e-ISBN : 978-629-98755-x-x

Design of the cover powered by
https://www.free-powerpoint-templates-design.com/