MONASH University

MONASH BUSINESS SCHOOL

Monash Business Analytics Major

Unit Summaries

Di Cook and Rob Hyndman

31 July 2024

The Master of Business Analytics synopsis is

> Become empowered as a quantitative citizen. By leveraging open data and the most powerful open source software available today you will learn how to fish, rather than be fed. This course is designed to develop thinking, computing and analytic skills for working with data for evidence-based solutions of today's problems. You will learn how to critically assess information provided by others by sourcing and analysing data yourself, with a strong emphasis on ethical and reproducible methods. This course is suitable for those with undergraduate degrees in quantitative disciplines including mathematics, statistics, computer science and engineering. The content is student-centred, inclusive, accessible and participants will be expected to operate as a connected cohort of like-minded people ready to build a better society.

and students completing this degree are expected to be able to:

1. Critical and creative scholars who:
    - produce innovative solutions to data analysis problems
    - apply research skills to business challenges
    - communicate effectively and perceptively
2. Responsible and effective global citizens who:
    - engage in an internationalised world
    - exhibit cross-cultural competence
    - demonstrate ethical values
3. effective data analysts able to:
    - identify and collect appropriate and relevant data
    - write computer scripts and programs to wrangle and plot data, fit models, make predictions and produce reproducible reports, presentations and web apps
    - interpret statistical models in the context of real data problems, and translate technical results into practical solutions.

Students must complete 12 core units and 4 elective units.

The following are core units and must be taken by all students:

- ETC5510 Data modelling and computing

- ETC5512 Wild-caught data
- ETC5513 Collaborative and reproducible practices
- ETC5242 Statistical thinking
- ETC5521 Diving deeply into data exploration
- ETC5523 Communicating with data
- ETC5550 Applied forecasting
- ETC5250 Introduction to machine learning
- ETC5543 Business analytics creative activity

And **two of the four units** from the following list,

- ETC5580 Advanced statistical modelling
- ETF5500 High dimensional data analysis
- ETC5450 Advanced R programming
- ETC5555 Statistical machine mining

And **four elective units** from the following list, or others by special request.

- ETC5410 Bayesian inference and data analysis
- ETF5248 Optimisation for business
- ETF5480 Decision modelling for business
- FIT5147 Data exploration and visualisation
- FIT5205 Data in society
- FIT5212 Data analysis for semi-structured data
- FIT9132 Introduction to databases
- FIT9136 Algorithms and programming foundations in Python
- MAT9004 Mathematical foundations for data science and AI

ETC5510: Introduction to data analysis

Unit synopsis

This unit introduces principles and techniques for exploring different types of data. You will learn concepts and practice in taking "data from the wild", reading different formats, tidying and wrangling it into shape for analysis, and creating useful visualisations to achieve effective data-driven decision-making. You will learn to use data to solve problems, and develop oral and written communication skills, using hands-on learning and live data analysis with code.

Learning outcomes

1. Learn to read different data formats, learn about tidy data and wrangling techniques.
2. Apply effective visualisation, descriptive analysis and modelling to understand relationships between variables, and make decisions with data.
3. Develop computing, communication, and reproducible reporting skills.

Topics covered include

- Workflow and introduction to R
- Tidy data
- Wrangling verbs/operations
- Visualising data
- Missing value handling
- Working with temporal and spatial variables
- Data collection: scraping, JSON
- Basic statistical models: form and fitting and diagnosis
- Basic computational models: algorithm (regression & classification trees), stopping, diagnostics
- Text analysis: wrangling text into data, sentiment tagging
- Basic algorithmic clustering

Textbook

Wickham, Çetinkaya-Rundel & Grolemund (2023). *R for Data Science*. 2nd edition, O'Reilly. https://r4ds.hadley.nz

ETC5512: Wild-caught data

Unit synopsis

Open data is a raw material for the digital age but unlike coal, timber or diamonds, it can be used by anyone and everyone at the same time. This unit will equip you with the skills to find, access, process, and prepare open data for further analyses. There are many ways that data can be collected, via experiments, observation, sampling, or sensors. There are many different formats, such as csv, JSON, spatial polygons, or html tables, to learn how to read. Accessing data might be through a database query, a GUI, an API, or scraping web sites. You will also learn about ethics and privacy issues with using open data, and how effectively curate your own open data.

Learning outcomes

1. Understand the definitions, allowed usage, digital identification and licensing of open data.
2. Know about common open data sources, how they are used and effectively search for new sources.
3. Explain the differences between data collection methods and the limitations for data analysis.
4. Work with the range of different data formats of open data, including APIs.
5. Understand ethical constraints and privacy limits when working with open data.
6. Recognise the components of effective curation needed for open data.

Topics covered include

- Open data: definitions, sources and examples
- Introduction to data collection methods
- Case study: US air traffic
- Case study: Australian census
- Case study: Australian election data
- Case study: Combining Australian census and election data
- Case study: Mortgage default
- Data ethics
- Data ethics and privacy
- Introduction to web scraping
- Case study: COVID-19 case data and the timing of lockdowns
- On the proper care and feeding of wild data

Textbook

See reading assignments on unit website: https://wcd.numbat.space

ETC5513: Collaborative and Reproducible Practices

Unit synopsis

This unit develops teamwork skills, through instruction on reproducibility and version control to conduct collaborative data analysis. These skills are practised in a team setting with a data analysis project.

Learning outcomes

1. Develop skills to create reproducible data analyses.
2. Understand the operation of a version control systems.
3. Utilise version control to integrate data analysis efforts of team members.
4. Effectively work with a group to construct a collaborative data story.

Topics covered include

- Introduction to collaborative and reproducible practices
- Reproducible reports using Quarto
- Introduction to version control systems: git and GitHub
- Reproducible reporting using Quarto, git and GitHub
- Deeper git knowledge, stashing and tools
- Reproducible reporting and version control systems
- Workflows for reproducible data analysis
- Reproducible reporting for specialised and broad audiences
- Advanced collaborative practices
- Reproducible workflows in consultancy

Textbook

See materials on unit website: https://rcp.numbat.space

ETC5242: Statistical thinking

Unit synopsis

This unit presents data analysis, statistical modelling and decision-making in the presence of uncertainty, using a computational approach. You will use different frameworks of probability that are useful for analysing real-world problems. Topics covered will include exploratory data analysis, simulation, permutation and randomisation methods, Bayesian analysis, decision theory, and model assessment and diagnosis.

Learning outcomes

1. Characterise and understand uncertainty using data.
2. Build statistical models to support decision-making, hypothesis testing and risk assessment.
3. Use randomisation methods in data collection and to assess causality and uncertainty.
4. Learn about and use concepts from probability.
5. Understand Bayesian and frequentist approaches to statistical modelling.
6. Further develop computational skills for statistical analysis.

Topics covered include

- Review of tidy data
- Estimation
- Hypothesis testing
- Resampling techniques, including the bootstrap
- Probability and probability distributions
- Maximum likelihood estimation
- Bayesian inference
- Decision theory
- Regression models, including multiple regression
- Model performance and diagnostics
- Re-sampling with regression models

Recommended reading

Introductory Statistics with Randomisation and Simulation, by D. M. Diez, C. D. Barr and M. Çetinkaya-Rundel, published by OpenIntro, 2014.

Modern Dive: Statistical Inference via Data Science, by C. Ismay and A. Y. Kim. Hardcopy published by CRC Press, 2020.

R for Data Science (2nd edition), by H. Wickham, M. Çetinkaya-Rundel and Garrett Grolemund. Hardcopy published by O'Reilly Media, Inc, 2023.

Statistical Thinking for the 21st Century, by R. A. Poldrack.

Doing Bayesian Data Analysis: A tutorial with R, JAGS and Stan, by J. K. Kruschke. 2nd edition, published by Elsevier, 2015.

ETC5521: Diving deeply into data exploration

Unit synopsis

Beyond modelling and prediction, data might have many more stories to tell. Exploring data to uncover patterns and structures, involves both numerical and visual techniques designed to reveal interesting information that may be unexpected. However, an analyst must be cautious not to over-interpret apparent patterns, and to use randomisation tools to assess whether the patterns are real or spurious.

Learning outcomes

1. Learn to use modern data exploration tools with many different types of contemporary data to uncover interesting structures, unusual relationships and anomalies.
2. Understand how to map out appropriate analyses for a given data set and description, define what we would expect to see in the data, and whether what we see is contrary to expectations.
3. Be able to compute null samples in order to test apparent patterns, and to interpret the results using computational methods for statistical inference.
4. Critically assess the strength and adequacy of data analysis.

Topics covered include

- Overview. Why this course? What is EDA?
- Learning from history
- Initial data analysis and model diagnostics: Model dependent exploration and how it differs from EDA
- Using computational tools to determine whether what is seen in the data can be assumed to apply more broadly
- Working with a single variable, making transformations, detecting outliers, using robust statistics
- Bivariate dependencies and relationships, transformations to linearise
- Making comparisons between groups and strata
- Going beyond two variables, exploring high dimensions
- Exploring data having a space and time context Part I
- Exploring data having a space and time context Part II
- Sculpting data using models, checking assumptions, co-dependency and performing diagnostics
- Extending beyond the data, what can and cannot be inferred more generally, given the data collection

Textbook

See the materials on the unit website https://ddde.numbat.space.

ETC5523: Communicating with data

Unit synopsis

This course teaches communication skills related to data analysis. You'll learn to build a website, a web app, reports and presentations as part of a data analysis workflow.

Learning outcomes

1. Effectively communicate data analysis, using a blog, reports and presentation.
2. Learn how to build a web app to provide an interactive data analysis.
3. Learn to construct a data story.

Topics covered include

- Basic communication theory and practice
- Narrative structure and reporting
- Captivating statistical presentations
- Data storytelling on the web
- Statistical model outputs and data tables
- Effective data visualisation
- Introduction to web technologies and styling
- R package and documentation
- Communicating data with interactive web apps
- Automated systems for code communication
- Case study: Communicating risk in a natural disaster

Textbook

See the materials on the unit website https://cwd.numbat.space.

ETC5250: Introduction to machine learning

Unit synopsis

This unit develops your ability to model multi-dimensional data using statistical and machine learning techniques. Topics covered include: dimension reduction with linear and nonlinear methods; supervised learning such as discriminant analysis, decision trees and forests, neural networks; and unsupervised learning such as k-means, hierarchical and model-based clustering. You will learn about conceptualising problems using the bias-variance trade-off and how to balance this when fitting models. Complex model fitting techniques will be covered including bagging, boosting, cross-validation, regularisation and constructing ensembles. An important component is learning how to diagnose your model, especially utilising high-dimensional visualisation methods, and explain your model with explainable artificial intelligence (XAI). You will develop practical skills in applying techniques to different problems using a suitable software environment that involves doing reproducible analyses.

Learning outcomes

1. Develop, select, and diagnose statistical and machine learning methods for supervised and unsupervised tasks.
2. Measure the uncertainty of a prediction or classification using resampling methods.
3. Efficiently conduct analysis tasks in a contemporary software environment.
4. Explain and interpret the analyses undertaken clearly and effectively.
5. Apply analytic tools to contemporary business problems.

Topics covered include

- Conceptual overview of ML, bias/variance, diagnostics, model choice, feature engineering, math/notation/computing
- Resampling: LOOCV, cross-validation, training/test
- Dimension reduction: PCA, PP, MDS, regularisation
- High-d vis: tours, par coords, mosaic plots
- Classification: logistic, LDA, trees, SVM, NN, deep learning
- Ensembles and boosting: forest, xgboost
- Explainability and interpretability
- Unsupervised classification: k-means, hierarchical clustering
- Model choice, diagnostics

Textbook

See the materials on the unit website https://iml.numbat.space.

- James, Witten, Hastie, & Taylor (2021). *An Introduction to Statistical Learning*. 2nd edition, Springer. https://www.statlearning.com
- Boehmke & Greenwell (2020) *Hands-On Machine Learning with R* https://bradleyboehmke.github.io/HOML/
- Kuhn and Silge (2023) *Tidy Modeling with R* https://www.tmwr.org

- Molnar (2023) *Interpretable Machine Learning* https://christophm.github.io/interpretable-ml-book/
- Hvitfeldt (2024) *Feature Engineering A-Z* https://feaz-book.com
- Cook & Laa (2024) *Interactively Exploring High-dimensional Data and Models in R* https://dicook.github.io/mulgar_book/

ETC5450: Advanced R programming

Unit synopsis

R is widely used as a tool for data analysis and one of the most popular programming languages. This unit delves into R from the programming aspect instead of using it as a data analysis tool. You will learn a variety of programming paradigms and delve into the foundations of R as a programming language.

Learning outcomes

1. Be familiar with the foundations of R programming.
2. Understand a variety of programming paradigms, including functional programming and object-oriented programming.
3. Utilise R's functional and object-oriented programming systems.
4. Integrate the concept of metaprogramming to evaluate and construct new functions

Topics covered include

- Foundations of R programming
- R package development
- Debugging and profiling
- Functional programming
- Object-oriented programming
- Object-oriented S3 and vctrs
- Reactive and literate programming
- Quarto and targets: efficient reproducible workflows
- Metaprogramming
- Interfacing with other languages
- Rewriting R code in C++

Textbook

See the materials at the unit website https://arp.numbat.space.

Advanced R, 2e by Hadley Wickham.

Advanced R Solutions by Malte Grosser, Henning Bumann & Hadley Wickham.

R packages, 2e by Hadley Wickham & Jenny Bryan.

Mastering Shiny by Hadley Wickham.

The {targets} R package user manual by Will Landau.

ETC5550: Applied forecasting

Unit synopsis

Business and economic data are collected over time against a backdrop of a changing environment. This unit introduces data wrangling and exploratory data analysis for such temporal context data. You will learn how to analyse, model and forecast time series data in diverse applications. Methods to be covered include the decomposition of time series, exponential smoothing methods, ARIMA modelling, and regression with auto-correlated disturbances. You will enhance your computational skills with exercises using R.

Learning outcomes

1. Understand common statistical methods used in business and economic forecasting.
2. Develop computer skills for forecasting business and economics time series data.
3. Provide insights into the problems of implementing and operating large scale forecasting systems.

Topics covered include

- Introduction to forecasting and R
- Time series graphics
- Time series decomposition
- The forecaster's toolbox
- Exponential smoothing
- Forecasting with ARIMA models
- Multiple regression and forecasting
- Dynamic regression

Textbook

See the materials at the unit website https://af.numbat.space.

Hyndman & Athansopoulos (2021). *Forecasting: principles and practice*. 3rd edition, OTexts. https://OTexts.com/fpp3/

ETC5555: Statistical machine learning

Unit synopsis

This unit covers the methods and practice of statistical machine learning for modern data analysis problems. Topics covered will include recommender systems, social networks, text mining, matrix decomposition and completion, and sparse multivariate methods. All computing will be conducted using the R programming language.

Learning outcomes

1. Identify and understand the statistical and computational trade-offs in modern data analysis problems.
2. Understand and apply machine learning algorithms to solve modern data analysis problems.

Topics covered include

- The learning model, hypothesis sets,
- Perceptron, Perceptron Learning Algorithm (PLA)
- Feasibility of learning, Hoefddings's inequality
- Linear models, Logistic regression, Cross-entropy
- (Stochastic) gradient descent
- Neural Networks, activation functions, backpropagation algorithm
- Deep neural networks (Keras to R interface)
- CNNs (Convolutional neural network)
- SVM, margin classifier, kernels
- Text mining, Vector space model, Bag of Words, N-grams
- Text and sequence learning, word embeddings, word2vec, Bias and fairness
- Implementation of most algorithms used in class from scratch using matrix operation (Linear models, gradient descent, Neural networks with matrices by hand and with code)
- Applied Machine Learning project as group work

ETC5580: Advanced statistical modeling

Unit synopsis

This unit introduces extensions of linear regression models for handling a wide variety of data analysis problems. Three extensions will be considered: generalised linear models for handling counts and binary data; mixed-effect models for handling data with a grouped or hierarchical structure; and non-parametric regression for handling non-linear relationships. All computing will be conducted using R.

Learning outcomes

1. Understand statistical models for handling common data analysis problems.
2. Develop skills for fitting, interpreting and assessing statistical models.
3. Develop computer skills for exploring and modelling different kinds of data.

Topics covered include

- Graphics for linear models and interpreting interactions
- Leverage, influence, and residual analysis for linear models
- Regression with binary responses
- Regression with binomial and proportional responses
- Regression with count responses
- Generalized linear models
- Random effects
- Mixed-effects models
- Nonparametric regression
- Nonparametric inference
- Additive models

Textbook

Faraway (2016) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 2nd edition, CRC Press. https://julianfaraway.github.io/faraway/ELM/

ETF5500: High-dimensional data analysis

Unit synopsis

In many fields of business, analysts must deal with data on many variables, for example, surveys with a large number of questions. In such cases, statistical tools known as multivariate methods must be used to analyse the data and drive business decisions.

This unit covers such methods in three sections: Cluster Analysis can be used to identify and predict differences between groups such as between distinct classes of customers or products; Principal Components Analysis, Correspondence Analysis and Multidimensional Scaling are dimension reduction methods that help analysts to visualise complicated datasets; and finally, Factor Analysis is used to explain and predict business outcomes.

Learning outcomes

1. Demonstrate an understanding of the role that multivariate statistical techniques such as factor analysis, categorical data analysis, cluster analysis, multidimensional scaling and correspondence analysis play in uncovering relationships and patterns in high dimensional data.
2. Appraise the strengths and limitations of these techniques.
3. Apply tools in R to analyse high dimensional data and generate solutions using the appropriate statistical techniques.
4. Demonstrate skills in using the appropriate statistical techniques from a user and provider perspective.
5. Demonstrate skills in communicating the results, including the visualization and reporting of the analysis, to aid decision making.

Topics covered include

- Distance and data
- Matrix algebra for multivariate analysis
- Cluster analysis
- Multidimensional scaling
- Principal components analysis
- Exploratory factor analysis
- Correspondence analysis

Main readings

Mardia, K. V., J.T. Kent, J.M. Bibby, (1979). *Multivariate Analysis*.

Izenman, A. J. (2008). *Modern multivariate statistical techniques. Regression, classification and manifold learning*

ETC5543:

Unit synopsis

This course involves successful completion of an internship or conduct an original research project.

Learning outcomes

1. Obtain experience working in a job or conducting research.
2. Apply analytics learning in a practical or research setting.
3. Effectively communicate analytics use in the work environment, or in the process of research.

ETC5410 Bayesian inference and data analysis

Unit synopsis

This unit introduces you to both foundational and methodological aspects of Bayesian inference and data analysis. Topics covered include a review of the philosophical and probabilistic foundations of Bayesian inference; the contrast between the Bayesian and frequentist (or classical) statistical paradigms; the use of prior information via the specification of objective, Jeffreys and subjective prior distributions; Bayesian linear regression; the use of simulation techniques in Bayesian inference, including Markov chain Monte Carlo algorithms; Bayesian analysis of Gaussian and non-Gaussian time series econometric models, including state space models; and the Kalman filter as a Bayesian updating rule.

Learning outcomes

1. Appreciate the importance of Bayesian statistical techniques and understand the differences between the Bayesian and frequentist statistical paradigms.
2. Acquire the skills necessary to derive Bayesian results analytically, in simple models.
3. Demonstrate an understanding of simulation methods and be able to implement these methods in empirically realistic models for data analysis.
4. Understand the Kalman filter and its role in Bayesian inference in state space models.

Topics covered include

- Introduction to Bayesian inference
- Conjugate and non-conjugate prior distributions
- Incorporating model uncertainty
- Markov chain Monte Carlo methods
- Applications and algorithms (using conjugate and non-conjugate priors) for
  - iid data from known parametric distributions
  - Linear regression models, with normal, Student-t or AR(1) errors
  - Gaussian state space models
  - Non-Gaussian state space models

Main readings

Greenberg (2013) Introduction to Bayesian Econometrics, 2nd edition, Cambridge University Press.

Casella and Berger (2002) Statistical Inference, 2nd edition, Duxbury Press.

Petris, Petrone and Campagnoli (2009) Dynamic Linear Models with R, Springer.

ETF5248 Optimisation for business

Unit synopsis

The ability to understand and mathematically formulate decision problems is a fundamental skill for managers in any organisation. This unit serves as an introduction to various optimisation techniques that are essential in business operations. You will learn to approach complex real-life problems, formulate appropriate models and compute solutions that offer managerial insights in various applications such as capacity planning, production management, and resource allocation. Methods include linear programming, integer programming and non-linear programming. Applications in management, marketing, accounting, finance and related fields are emphasised.

Learning outcomes

1. Select the appropriate basic optimisation models according to the business context.
2. Formulate the optimisation problem for management decision making.
3. Identify the potential limitations of the models and suggest creative solutions to overcome model weaknesses.
4. Provide interpretations and managerial insights according to the solution of the optimisation problem.
5. Implement and solve the optimisation model using scalable software.

Topics covered include

- LP basics and examples
- Duality and sensitivity
- Network flow problems
- (Mixed) Integer programming
- Advanced LP reformulation

Main readings

D. Bertsimas and J. N. Tsitsiklis, Introduction to Linear Optimization.

W. L. Winston, Operations Research: Applications and Algorithms.

B. Taylor, Introduction to Management Science: Global Edition.

R. J. Vanderbei, Linear Programming: Foundations and Extensions.

ETF5480 Decision modelling for business

Unit synopsis

This unit is designed to develop quantitative problem solving skills. The emphasis is on recognising and identifying, modelling and solving problems arising in business-decision. Various optimisation techniques used in the business decision-making process will be introduced. Topics include benchmarking with data envelopment analysis, game theory, queuing models, Markov chains and simulation modelling. Applications in business, particularly in management, marketing, banking and finance will be used in the demonstration.

Learning outcomes

1. Understand the complementary nature of the rational and behavioural approaches to decision making.
2. Select and apply quantitative modelling concepts to problems arising from business.
3. Formulate an appropriate model to analyse various business decision problems.
4. Apply tools available in Microsoft Excel and algorithms to generate solutions and appraise the solutions through the use of sensitivity analysis.
5. Communicate the analysis and results of business decision problems.

Topics covered include

- Queuing models
- Simulation
- Decision tree methods
- Game theory: LP and duality recap
- Data envelopment analysis: linear fractional programming
- Optimization in data analytics: convex optimization and quadratic programming

Main readings

D. Bertsimas and R. M. Freund, Data, Models, and Decisions: The Fundamentals of Management Science.

S. J. Wright, B. Recht, Optimization for Data Analysis.

R. J. Vanderbei, Linear Programming: Foundations and Extensions.

S. A. Klugman, H. H. Panjer and G.E. Willmot, Loss Models: From Data to Decisions.

D. Bertsimas and J. Dunn, Machine Learning Under a Modern Optimization Lens.

FIT5147 Data exploration and visualisation

Unit synopsis

This unit introduces statistical and visualisation techniques for the exploratory analysis of data. It will cover the role of data visualisation in data science and its limitations. Visualisation of qualitative, quantitative, temporal and spatial data will be presented. What makes an effective data visualisation, interactive data visualisation, and creating data visualisations with R and other tools will also be presented.

Learning outcomes

1. Perform exploratory data analysis using a range of visualisation tools.
2. Describe the role of data visualisation in data science and its limitations.
3. Critically evaluate and interpret a data visualisation.
4. Distinguish standard visualisations for qualitative, quantitative, temporal and spatial data.
5. Choose an appropriate data visualisation.
6. Implement static and interactive data visualisations using R and other tools.

FIT5205 Data in society

Unit synopsis

In the digital age, how we communicate, conduct business and socialize is revolutionizing our world. The explosion of data-driven and artificial intelligence technologies are transforming the way we live and work, altering relationships between government and citizens, businesses and consumers, researchers and the researched, public and private sectors, and individuals, communities and societies.

While data driven and algorithmic technologies can potentially help grow the economy, improve health and education, support national security, protect the environment, enable more energy efficiency, drive innovation and progress, and support more resilient, sustainable communities and cultures, there is a growing need to identify and address the risks associated with their design, development and use. In this unit the problems associated with the application of these information technologies in government and big business that leads to greater surveillance of citizens by states and consumers by businesses, disempowerment of individuals and vulnerable communities, increased discrimination, threats to social inclusion, social justice, human and civil rights, self-determination and privacy, and widens the divide between the data haves and have-nots will be explored, alongside the emerging ethical, legislative and strategic frameworks that aim to address these concerns.

Learning outcomes

1. Describe and analyse digital media/data ecosystems.
2. Analyse issues relating to the benefits and risks of big data in society.
3. Explain key ethical design principles and their relevance to the design and development of data-driven and algorithmic technologies.
4. Investigate how rights in data, rights to self determination, privacy rights, access rights, discovery rights, IP and copyright apply in different contexts/scenarios.
5. Develop recommendations for socio-legal and policy frameworks and strategies for transparency, good governance, accountability and ethical practice in data management and use, including data rights management.
6. Analyse needs and issues relating to the use big data to support resilient, sustainable communities and cultures.

FIT5212 Data analysis for semi-structured data

Unit synopsis

Semi-structured data is one of the fastest growing kinds of data in both the public and private sector, for instance in health. Email collections with sender-recipient graphs, metadata and text content is one example. This unit will explore basic forms of semi-structured data: text, time-sequence data, graphs and multiple relations in a database. Basic machine learning algorithms for these kinds of data will be analysed and applied. Some characteristic industry problems for the application of semi-structured data will also be investigated.

Learning outcomes

1. Appraise what kinds of semi-structured data exist and the problems they present for analysis.
2. Analyse different kinds of algorithms for different kinds of semi-structured data.
3. Develop and modify some standard algorithms for semi-structured data.
4. Examine some characteristic industry problems involving semi-structured data, and analyse the suitability of different algorithms.

FIT9132 Introduction to databases

Unit synopsis

This unit will introduce the concept of data management in an organisation through relational database technology. Theoretical foundation of relational model, analysis and design, implementation of relational database using SQL will be covered.

Learning outcomes

1. Apply the theories of the relational database model.
2. Develop a sound relational database design.
3. Implement a relational database based on a sound database design.
4. Manage data that meets user requirements, including queries and transactions.
5. Contrast the differences between non-relational database models and the relational database model.

FIT9136 Algorithms and programming foundations in Python

Unit synopsis

This unit introduces the Python programming and the basics of data structure and algorithms including their design, analysis and implementation in Python.

Students will experience working with Python implementation of data structures and algorithms widely used in modern programming language to solve simple problems. Topics covered in this unit are the programming basics including IOs, control structures, and concepts of object-oriented programming; data structures and algorithms including lists, stacks, queues, trees, recursion and searching/sorting algorithms.

Learning outcomes

1. Apply best practice Python programming constructs for solving computational problems.
2. Restructure a computational program into manageable units of modules and classes using the object-oriented methodology.
3. Demonstrate Input/Output strategies in a Python application and apply appropriate testing and exception handling techniques.
4. Investigate useful Python packages for scientific computing and data analysis.
5. Experiment with data manipulation, analysis, and visualisation technique to formulate business insight.

MAT9004 Mathematical foundations for data science and AI

Unit synopsis

Mathematical topics fundamental to computing and statistics including trees and other graphs, counting in combinatorics, principles of elementary probability theory, linear algebra, and fundamental concepts of calculus in one and several variables.

Learning outcomes

1. Use trees and graphs to solve problems in computer science.
2. Apply counting principles in combinatorics.
3. Describe the principles of elementary probability theory, evaluate conditional probabilities and use Bayes' Theorem.
4. Demonstrate basic knowledge and skills of linear algebra, including the manipulation of matrices, solution of linear systems, and evaluate and apply determinants.
5. Explain fundamental concepts in calculus including basic differentiation and integration, and composite, inverse and parametric functions.
6. Perform key skills in the calculus of functions of several variables including the calculation of partial derivatives, find tangent planes and identify stationary points, root findings and convexity for optimisation.