MIS772 Predictive Analytics

T11: Review and Exam

Fundamental Processes of Predictive Analytics

- Translating a Business Problem into (a) specific class(es) of analytical problem(s)
- Data Sources, Preparation
- Explorative Analytics that inform Predictive Modelling
- Predictive model development, evaluation, optimisation
- Returning back to the business problem, addressing original business question(s)
- Deployment of models

Classes of problems studied

- **Classification:** predicting which class data belongs to? A or B?
- **Regression (estimation):** predicting how much or how many
- **Text mining:** finding patterns in unstructured textual data
- **Clustering (segmentation**): detecting groups that data belong to
- **Market Basket Analysis:** what goes together with what
- **Timeseries Forecasting:** predicting future trends
- **Anomaly detection:** which values do not belong

How do we know if our modelling is good or if they can be optimised?

We look at relevant performance indicators for different classes of problems

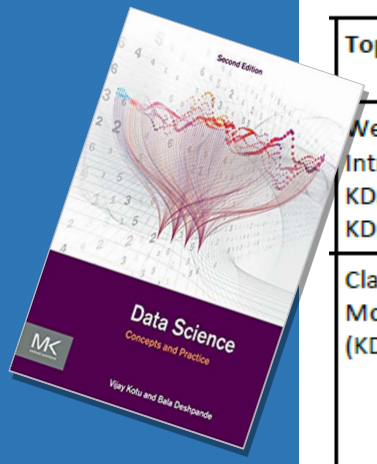
E.g., accuracy, kappa in Classification;

RMSE, R^2 in Regression

Davies Boudin in Clustering

Etc.

DEAKIN UNIVERSITY



Topic (Refs)	Special learning activities (with RM Studio)
Welcome and Introduction (KD1-KD3, KD15)	Data attributes, data acquisition and exploration, data visualisation, analytics process, model training, validation and testing
Classification Models (KD4.1-KD4.4)	Classifiers (k-NN, decision trees, logistic regression, support vector machines) binomial and multinomial classification, hold-out validation, performance measurements (accuracy, recall and precision), confusion matrix
Advanced Data Classification (KD4.7, KD8, KD15.5-KD15.6)	Cross-validation (k-fold, bootstrap, and LOOCV), class imbalance, accuracy vs kappa, other performance measurements (ROC and AUC), overfitting and underfitting, bias and variance, ensembles (voting, random forests, bagging, stacking and boosting), model optimisation
Introduction to Text Mining (KD9)	Text representation, Text mining, dimensionality reduction, introduction to sentiment analysis, text-based predictive models

Topic (Refs)	Special learning activities (with RM Studio)
Estimation Models and Multiple Linear Regression (KD5.1)	Multiple regression and other estimation methods, model diagnostics (coefficients, t-statistics, p-values, f-ratio), variables selection and regularisation, dealing with multi-collinearity (correlation, VIF and tolerance), dealing with missing values (replacement and imputation), validation of linear models, (linear and non-linear), model optimisation
Data Clustering & Segmentation Analysis (KD7)	k-Means and other types of clustering, segmentation, optimisation of clustering, Principal Component Analysis (PCA), cluster visualisation with PCA
Market Basket Analysis (KD6)	Association rule mining and sequential pattern analysis, Rule generation algorithms, Evaluation metrics
Anomaly Detection (KD13)	Outliers vs anomalies, distance, density, global and local anomaly detection, Singular Value Decomposition (SVD), anomaly visualisation with PCA and SVD
Time Series Analysis and Forecasting (KD12)	Time series analysis, single and multi-horizon forecasting, stationarity and smoothing, auto-correlation, autoregressive models, motifs, shapelets and discords
Model Application and Deployment (Readings)	Wrapping up, model application (reflection on previously practised techniques), analytics process deployment, platforms

All topics are examinable

Including all readings, Lectures, seminars, assignment tasks, and online resources

DEAKIN UNIVERSITY

Preparation for Exam

- *Revise your assignment work* based on *provided feedback*.
- *Address assignments problems* as reflected in the assessor's feedback and your reflection.
- Ensure you have completed all *seminar exercises*.
- Study lecture notes.
- *If you still have any doubts* about some aspects of the lectures, workshops or assignments
  - Read the most relevant *textbook chapters*
  - *Research*
- We expect that you have helped yourself before you ask for help!

- Study the *sample exam*
  - Try answering the questions
  - Compare your answers and problem solutions from the what we've learned throughout the trimester and do a bit of research where needed

- *Remember that this is also the busiest time for teaching staff.*
  - They will also be involved in marking assignments for many units at the time of your exam preparation.
  - Often they use this time to conduct their research, and prepare for future teaching.

- Please use CloudDeakin to ask questions rather than sending emails.

DEAKIN UNIVERSITY

Structure of Exam

- *This exam is in the form of an online quiz.*
- It consists of 120 marks which contribute to 50% of the total assessment in this unit.
- This unit has a hurdle requirement. You need to achieve at least 50% of the marks available on the examination.
- Please read the document with important exam pre-information posted on the unit website!

- *A single business case study* will support answers to all questions.
- *One concepts question* will cover knowledge of ideas, concepts and methods, it will require analytic answers, such as:
  - explain,
  - compare and contrast,
  - describe similarities and differences, etc.
- *Four problem-solving questions* covering three modules of teaching, e.g.
  - Create an analytic process using model X to solve this problem
  - Explain the given process
  - Explain how to improve the provided model X
  - Explain the results returned by the model X
  - Explain the performance of model X as shown in the provided charts and statistics
- *Answers need to be in the requested form*, e.g. “in a paragraph” or “in a point form”.
- *All answers must be clear and precise* and be provided in the text boxes in the online quiz.

DEAKIN UNIVERSITY

Sample Exam Case Study

Not a real-exam case

The following is a mini-case study.

Travelairex is reviewing various aspects of passengers' air travel experience across the globe. Travelairex are interested in passenger recommendation in the order of their importance: (1) airlines, (2) airports, (3) lounges and (4) seats. Currently, the passenger reviews are collected by Travelairex to include a short survey and some descriptive comments about various aspects of their travel. However, in the future Travelairex would like to collect such information from social media, such as Twitter and Facebook, where the passengers would not be aware of their monitoring and opinion probing.

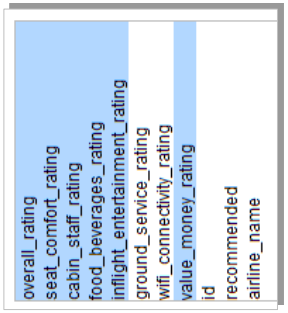
Your team's objective is to conduct a segmentation analysis of data, then create an estimation of the passenger's overall rating as well as recommendation, using structured (non-text) data only and then text data only. Create a model ensemble.

Wikimedia: Wongm

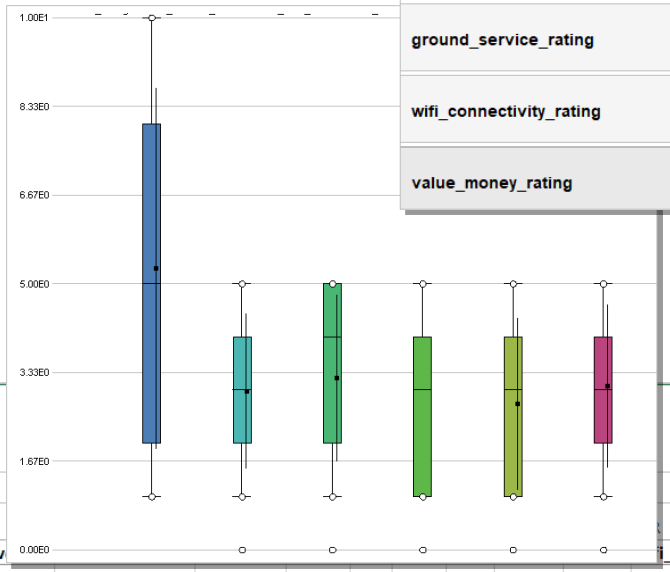
Data

Examples: 41396 / Attributes:

- (nominal) airline_name: 41396
- (nominal) link: 41396
- (nominal) title: 41396
- (nominal) author: 41396
- (nominal) author_country: 39805
- (date) date: 41396
- (text) content: 41396
- (nominal) aircraft: 1278
- (nominal) type_traveller: 2378
- (nominal) cabin_flown: 38520
- (nominal) route: 2341
- (integer) overall_rating: 36861
- (integer) seat_comfort_rating: 33706
- (integer) cabin_staff_rating: 33708
- (integer) food_beverages_rating: 33264
- (integer) inflight_entertainment_rating: 31114
- (integer) ground_service_rating: 2203
- (integer) wifi_connectivity_rating: 565
- (integer) value_money_rating: 39723
- (binomial) recommended: 41396



Name	Type	Missing
Cluster airline_name	Polynominal	0
overall_rating	Real	0
seat_comfort_rating	Real	0
cabin_staff_rating	Real	0
food_beverages_rating	Real	15
inflight_entertainment_rating	Real	0
ground_service_rating	Real	9736
wifi_connectivity_rating	Real	10111
value_money_rating	Real	0



airline_clean.xlsx - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW TEAM

Jacob Cybulski

A1 airline_name

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	airline_name	link	title	author	author_country	content	date	aircraft	type_trav										value_	recommended
50	adria-airways	/airline-re	Adria Airw	David Grey		Round trip LGW - LJU in Economy Class bc	15/09/2005													0
51	adria-airways	/airline-re	Adria Airw	Markus Lattner		I have used their VIE-FRA-VIE service sev	3/12/2004													0
52	adria-airways	/airline-re	Adria Airw	Alex Stare		Flew march 12/14 on JP from ZRH-LJU-ZRI	18/03/2004													0
53	aegean-airlines	/airline-re	Aegean Ai	N Sunder	United States	Brand new A320, interior spotless. Crew	21/07/2015	A320	Solo Leisure	Economy	Mykonos to	8	4	4			4	4		1
54	aegean-airlines	/airline-re	Aegean Ai	Dimitrios	Greece	I booked a premium economy ticket so I	21/07/2015	Dash 8 Q4	Solo Leisure	Economy	Athens to S	8	4	5	4		2	3		1
55	aegean-airlines	/airline-re	Aegean Ai	Kim Buhrr	United States	2 flights - HER to ATH (short 50 minutes) a	12/07/2015		FamilyLeisure	Economy	ATH to MXP	10	4	5	4		5	5		1
56	adria-airways	/airline-re	Adria Airw	Alan McLean		Round trip LGW - LJU in Business Class. Cl	1/11/2003													0
57	adria-airways	/airline-re	Adria Airw	Andrew Proud		In July 2003 I flew London (Gatwick)-Ljub	31/07/2003													0
58	adria-airways	/airline-re	Adria Airw	G Chew		I have just returned from a long weekend	6/12/2002													0
59	adria-airways	/airline-re	Adria Airw	Francisco	Camino	I have flown Adria recently on a LH code :	25/08/2002													0

Final Advice

Train hard and win easily!

And good luck will not be needed…

Wishing you the very best for your future career successes

from the MIS772 teaching team:

Arman, Kaushi, Joerin and Mina

Please Evaluate the MIS 772 team!

And stay in touch!

DEAKIN UNIVERSITY