[1]

Witten, I., Frank, E., Hall, M., Pal, C*.* “Data Mining: Practical Machine Learning Tools and Techniques”. *https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+algorithms+data+mining&ots=8HIMsgnCy8&sig=0IRitlvtXZ-Wx7uSkRyff1y4W6o#v=onepage&q=probabilistic%20methods&f=false*

* Book on data mining and machine learning
* Contains information on basic algorithms
  + Probabilistic
  + Trees
  + Rules
  + Linear models
  + Clustering
* Credibility and evaluation information
* Information on advanced ML techniques
  + Trees and rules
  + Extending instance based and linear models
  + Data transformations
  + Probabilistic methods
  + Deep learning
  + Beyond supervised and unsupervised
  + Ensemble learning
  + Applications beyond

SUMMARY

[2]

Data mining for education

http://users.wpi.edu/~rsbaker/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf

* Educational data mining (EDM) – area of scientific inquiry around the development of methods for making discoveries within the unique kinds of data that come from educational settings
* In 2008, establishment of the annual international conference on EDM
* EDM has advantages over traditional educational research paradigms (lab experiments, design research)
  + Educational data repositories provide a base which makes EDM highly feasible
    - Ecologically valid (data is about performance and learning of genuine students, in genuine settings)
    - This is often a challenge for researchers in other paradigms
* **Prediction** 
  + Important to consider the degree to which these labels may in fact be approximate, or incompletely reliable
  + Prediction methods can be used to determine important features for prediction
  + Can also be used to predict what the output value would be in contexts where it is not desirable to directly obtain a label for that construct
  + Broadly three types of prediction: classification, regression and density estimation
* Relationship mining
  + Goal is to discover relationships between variables
  + Broadly, 4 types: association rule mining, correlation mining, sequential pattern mining, casual data mining
  + Relationships found must satisfy 2 criteria: statistical significance and interestingness
* Information on discovery with models and distillation of data for human judgment (visualization)
* Applications
  + Improving student models providing detailed info about student (characteristics/states)
  + Discovering/improving models of the knowledge structure of the domain
  + Studying the pedagogical support provided by learning software
  + Scientific discovery about learning and learners

[3]

Predicting grades in college courses

*http://files.eric.ed.gov/fulltext/EJ829428.pdf*

* Uses high school grade point averages (HSGPA) and SAT scores to predict cumulative grades in different college courses
* Results displayed a trend of number of students doing well at a high level (GPS 3.5+) increases for students as SAT scores increase for students with similar high school grades
* Used multiple regression equations

[4]

Using data mining for predicting relationships between online question theme and final grade

*file:///Users/ethankenwrick/Downloads/Using\_Data\_Mining\_for\_Predicting\_Relationships\_bet.pdf*

* Uses EDM and regression analysis to analyse live video streaming students’ online behaviours and performance
  + Measures login frequency, participation, number of chat messages and questions students submit were analysed, along with final grades
* Results showed considerable variability in students’ questions and messages
* Unlike previous studies, suggests no correlation between all measures and a students success
* Does reveal that combining EDM with traditional statistical analysis provides a strong and coherent analytical framework, capable of enabling a deeper understanding of students’ learning behaviours and experiences

[5]