University of South Wales
Prifysgol De Cymru

# MS4S08 - Applied Statistics for Data Science

Penny Holborn
Penny.Holborn@southwales.ac.uk

University of South Wales
Prifysgol De Cymru

# Module Aims

- To provide students with an understanding of the core statistical analysis required for Data Science.

- To provide students with the practical knowledge of the statistical techniques used in manipulating and managing complex datasets such that they are able to assess practical situations and interpret real-world applications.

University of South Wales
Prifysgol De Cymru

# Module Plan

| Week | Topics | Staff |
|---|---|---|
| 1 | Intro to SAS Online<br>Summary statistics<br>Basic Probability | PH |
| 2 | Data Visualisation<br>Exploratory Data Analysis<br>Distribution analysis | PH |
| 3 | Hypothesis testing<br>Evaluation metrics<br>Reporting | PH |
| 4 | Tests for means/medians<br>Confidence intervals | PH |
| 5 | Correlation<br>Linear Regression<br>Residuals | AP |
| 6 | Logistic Regression<br>Interpretation of models | AP |
| 7 | Multivariate Analysis – PCA | AP |
| 8 | Multivariate Analysis – FA/Cluster Analysis | AP |

University of South Wales
Prifysgol De Cymru

# Assessments

| | Description | Hand-in date |
|---|---|---|
| **Assessment 1 50%** | Collect, analyse and interpret data and present results through a variety of visualisations. | **24/11/2020** |
| **Assessment 2 50%** | Perform a range of statistical analysis using software and formally report results. | **12/01/2021** |

University of South Wales
Prifysgol De Cymru

# Big Data

## 3V's of BIG DATA

Immense volume of data

- Each day we create 2.5 quintillion bytes


The velocity at which data is streaming

- Moore's Law – last 50 years computing
- Power x2 every 18 months


Huge variation in data

- Structured data – organised (rows,columns)
- Unstructured data – unorgnaised (text)

University of South Wales
Prifysgol De Cymru

# Why is Big Data important?


- Market basket analysis - popular purchases, insight into which products to promote.
- Customer churn - predict customer losses to a competitor.
- Fraud detection - single out potentially fraudulent transactions.
- Direct marketing – target individuals.
- Interactive marketing - internet can provide customer insight.
- Trend analysis - analyse shopping habits, compare monthly purchases.
- Optimisation - re-calculate delivery vehicles routes whilst they are on the road.



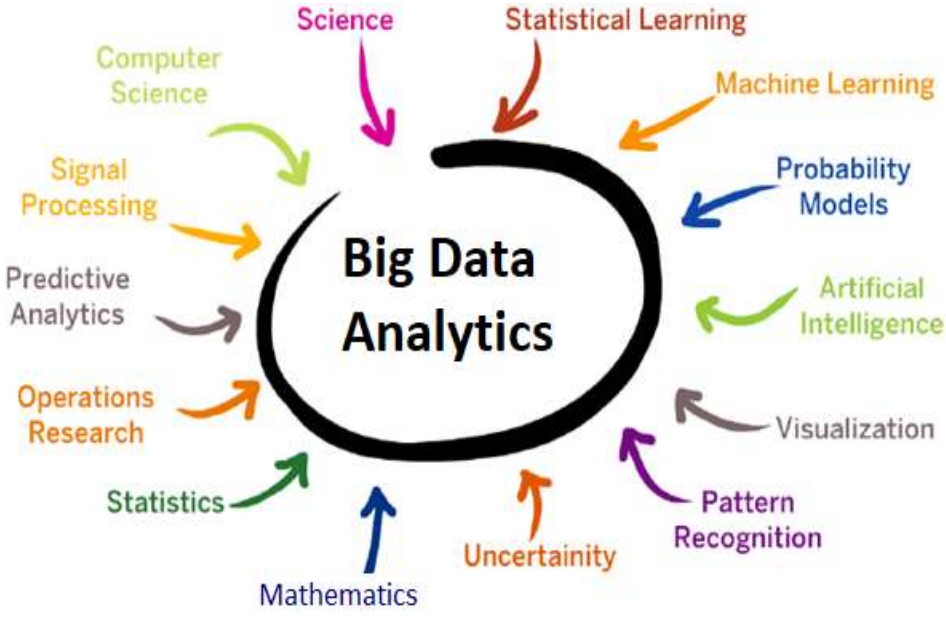
University of South Wales
Prifysgol De Cymru

# Big Data Analytics – Data Science

- The techniques and processes to manage, manipulate and analyse large data sets to discover patterns and other useful information.

- Big Data Analytics can provide businesses and organisations with insight to make more-informed decision.

- Methods at the intersection of Machine Learning, Statistics, Operational Research and Artificial Intelligence, etc.



University of South Wales
Prifysgol De Cymru

# Contact


Contact Details
Dr Penny Holborn
Faculty of Computing, Engineering and Science
School of Computing & Mathematics
Senior Lecturer
penny.holborn@southwales.ac.uk
(01443) 6 54370
Treforest Campus
@pennyholborn