NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Artificial Intelligence & Analytics: Competing in the Age of Big Data

Introduction

With the explosion of **Big Data**, **Artificial intelligence (AI)** and **Data Science (DS)** have become immensely popular across various fields, including business, education, and academia. AI and DS are revolutionizing the way businesses operate by enhancing efficiency, improving decision-making, and enabling the creation of innovative products and services.

A key component of modern AI and DS is learning**, known respectively as Machine learning (ML) or Statistical learning (SL)**, which is to formulate **learning** from the vast amount of **information**. Indeed there is a lot of cross fertilization from this two fields. We will address key questions such as: *How to discover patterns and relationships in data? How to extract useful information and knowledge? How to extract insights and make informed business decisions? How to automate learning?*

Today, AI technology runs an increasing part of everyday life, often without you knowing it. Modern AI/ML applications include advanced web search engines (e.g. Baidu, Bing, Google, etc), recommendation systems (e.g. Amazon, Netflix, Tiktok, Youtube, etc), human speech recognition (e.g. Alexa, Siri, Xiao AI, etc), self-driving cars (e.g., Tesla), strategic game systems (e.g. AlphaGo), protein folding (e.g. Alphafold), and many more. Companies in almost every industry are focused on exploiting data for competitive advantage and people with AI and DS background are highly sought after.

Course Objectives

In this course, we will take a holistic approach to understand the key factors involved, from data collection, analysis to prediction and insight. We will learn widely used methods for describing and summarizing data, discover patterns and trends in data, as well as methods and techniques in analyzing and modelling real world data to solve business problems. Emphasis will be on merging technical skills with critical thinking to ensure that robust data science pipelines are being created for business benefit. We will study case studies on market leaders and innovative

startups and explore how companies leverage massive amounts of data and sophisticated analytics to succeed in today's data-driven environments.

The course provides a view of AI and DS that focuses on ideas, models, and principles as well as discussion of some general scenarios and applications.

- Can you understand the basics of data and analytics?
- Can you analyze data or to consume it effectively?
- How are companies putting modern analytical technologies in place?
- How are businesses using data in their decision-making?
- How do you build models to learn from data?

Python has become a standard language in data science and with a wide array of resources available online, e.g., Kaggle, Github, Google ML, etc. In this course, we will utilize python and its array of useful libraries to analyze and model data.

- To understand the core concepts of Python programming language.
- To use the computational features in the Python programming language.
- To read in and analyze data in Python environment.
- To understand how data can be visualized using Python visualization libraries.
- How to model data using different techniques in supervised and unsupervised learning such as linear regression, logistic regression, k-nearest neighbor clustering, support vector machines, etc.

We will study case studies on market leaders and innovative startups and explore how companies leverage massive amounts of data and sophisticated analytics to succeed in today's data-driven environments. Given the rapid development of DS, the course will focus on providing you with the skills necessary to independently learn techniques not covered by the course, in order to address the more complex problems that you may encounter in the future.

Course outline

Session 1: An introduction to AI and DS: *concepts and applications.*

- What is AI and DS?
- How AI is revolutionizing the world
- Common AI applications
- Generative AI and evolution of large language models
- Top Trends in AI and DS
- Python, Anaconda, and libraries

Session 2: Descriptive Analytics – statistics and visualization

- From Business problems to Data Mining tasks
- Descriptive Statistical Measures
- The Value of Data Visualization
- Python: pandas, matplotlib and seaborn
- Case study: Titanic dataset

Session 3: Introduction to Predictive Modeling

- Models, Induction, and Prediction
- Modeling Relationships and Trends in Data
- Supervised vs Unsupervised learning
- Finding "optimal" model parameters based on data
- Optimization, gradient descent, and applications
- Python: scikit-Learn

Session 4: Decision Analytic Thinking: What Is a Good Model?

- Overfitting and Its Avoidance
- train/dev/test sets, optimization, regularization
- Similarity, Neighbors, and Clusters
- Expected value as a key evaluation framework
- Case study: Iris, dataset

Session 5: Group Presentation by Participants

Duration

3 hours x 5 sessions= 15 hours

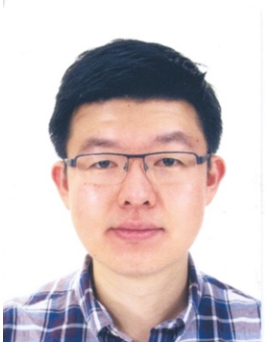
Teaching Methodology

General: Because class attendance and success in class are positively correlated, students are expected to attend lectures punctually and to participate actively in class.

Pedagogy: In this age of internet and social media, the use of interactive tools to make learning a more active process will become the new norm in teaching. The course will use multimedia (e.g., YouTube clips), online tools to make lectures more lively, engaging, and

interactive. They are powerful visual and experiential stimuli that help to create a more immersive and value-added learning environment. Oftentimes, students become inspired to learn more after watching YouTube clips about real-life applications of course content.

About the Instructor



Dr Yong Ee Hou is an Assistant Professor from the School of Physical and Mathematical Sciences at Nanyang Technological University. He received his BSc in Mathematics, BSc in Physics, MSc in Statistics from Stanford University, and PhD in Physics from Harvard University. His research focuses on using new and emerging state-of-the-art methods from mathematics, physics, artificial intelligence, machine learning, topological data analysis, and statistics to analyze big data of complex systems such as DNA, RNA, biological networks, social networks, financial markets, etc., in order to understand their behavior, pattern and interaction. His work has been published in top journals such as Science, PNAS, Physical Review Letters, etc.