

ANL488 Project List for Jul 2022 Semester

No	Project Title	Description
1	Modelling and Forecasting Time Series (Supervisor: Dr Karl Wu)	<p>This project focuses on modelling and forecasting time series using preferably ARIMA models (or other techniques that we have learned in ANL317 Business Forecasting). We will conduct a complete time series analysis including examining the properties of the series such as stationarity, seasonality, the periodogram (ACF and PACF), the goodness of fit of the estimated ARIMA models as well as the residuals. Eventually, we will try to obtain a reliable prediction of the future outcomes of the series. If we discover volatility in the time series, we may also try to implement some more challenging models such as the GARCH/ARCH as a complementary element to the conventional ARIMA approach.</p> <p>The time series we are going to work with should be univariate, meaning that it will be a series of a single variable. The series can be either from the fields of social science (e.g. studies on education, observations of political issues), economics (e.g. consumer price index), finance (e.g. stock prices), medicine (e.g. epidemic study) etc. You are also most welcome to suggest a certain field or topic where we can find time series for forecasting.</p> <p>We will be using either SAS Forecast Studio and/or R for this project. Those interested please email Dr. Wu at karlwuky@suss.edu.sg by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
2	Analysing zero-inflated data with R (Supervisor: Dr Karl Wu)	<p>Zero-inflation occurs when there are more zeros observed in a variable of interest than expected. Zero-inflated data are e.g. daily precipitation data (there are days in which 0mm rainfall are recorded), household income data (data could be zero-inflated if the unemployment rate during a certain period is high), number of defective items in production batches (some machines take longer to produce defective items than others). Zero-inflation is not a desirable feature since we cannot use traditional statistical approaches for our analysis. The normality assumption is usually violated here. In fact, zero-inflated data follow their own probability distribution. Analysing this type of data requires specific models which are integrated in most of the software nowadays.</p> <p>This project will focus on using R programming language to model zero-inflated data. The R package “zeroinfl” is specifically designed for modelling zero-inflated count data. Depending on the dataset to be analysed eventually, interested students may also try to develop their own R code for existing zero-inflation models. Hence, students should be preferably interested in mathematics/statistics and learning/using R. Those interested are to email Dr Wu at karlwuky@suss.edu.sg by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
3	Data Mining using open datasets (Supervisor: Dr Liu Wenting)	<p>You may propose a possible analytics application in an area of work that you are sufficiently familiar with (business or non-business related). You can use publicly available datasets from open data sources or fictitious data to complete your work. You need to distil a concrete description of the business situation, understand the significant opportunities and challenges based on facts and data.</p>

		<p>You may propose one or more modeling technique to analyze the datasets and generate insights.</p> <p>Examples of open data sources (include but are not limited to):</p> <ol style="list-style-type: none"> 1. https://data.gov.sg/ 2. https://datamall.lta.gov.sg/content/datamall/en.html 3. World Bank Open Data (https://data.worldbank.org/) 4. https://www.kaggle.com/datasets <p>Students who are strong/comfortable with Python/R programming are to send their CV to wentingliu@suss.edu.sg and cc (leeyh@suss.edu.sg) by noon 6 Jun 2022.</p>
4	<p>Advanced analytics to assess Corporate Risk and predict default</p> <p>(Supervisor: Mr Chua Poh Chai)</p>	<p>Traditional Corporate Risk models are too slow to respond to the changing business environment. By the time, the distress corporates are identified and downgraded, they are almost in default. This results in substantial credit losses for the banks and financial institutions.</p> <p>This project aims to develop advanced analytical models which are more responsive to market conditions and can predict distress/default in a more-timely manner so that banks and financial institutions can respond much earlier and reduce credit losses.</p> <p>This project will investigate various machine learning techniques, including deep learning, to build predictive corporate risk models. More importantly, it will break down the predictive model metrics into intelligible terms which practitioners can understand so as to achieve buy-in and deployment in banks and financial institutions.</p> <p>With reference to the Monetary Authority of Singapore (MAS) Financial Stability Review (FSR) 2018, P73-77, data can be obtained from providers like Bloomberg and Refinitiv (for publicly traded companies), Accounting and Corporate Regulatory Authority of Singapore (ACRA) and Ministry of Law.</p> <p>Students who are keen in this project can email Chua Poh Chai at pcchua002@suss.edu.sg by noon 6 Jun 2022 with their CV and copy leeyh@suss.edu.sg. As this project focuses on financial analytics, students' knowledge in finance, statistics and machine learning, together with strong Python/R programming, will be helpful.</p>
5	<p>Text/Sentiment Analysis</p> <p>(Supervisor: Mr Chua Poh Chai)</p>	<p>This project aims to extract meaningful insights to address business problems. This could be the reviews of products and services offered by business, in-depth understanding of targets (eg consumers, patients, students etc.) or sentiment on the ground. For example, airlines/telcos, healthcare, education or public sector agencies. With insights, these organisations can improve their offerings to achieve business goals.</p> <p>For sentiment analysis, predictive models may be developed.</p>

		<p>Students will source for the data for this project.</p> <p>As this project focuses on text mining and/or sentiment analysis, students are expected to have completed ANL303 and ANL312.</p> <p>Students who are keen in this project can email Chua Poh Chai at pcchua002@suss.edu.sg by noon 6 Jun 2022 and copy leeyh@suss.edu.sg.</p>
6	<p>Learning Analytics: Predictive Model on at-risk students</p> <p>(Supervisor: Mr Edwin Seng)</p>	<p>Research on the use and application of learning analytics and how learning analytics can be implemented in the higher education sector, and its expected benefits and interventions. Through preliminary research, learning analytics can facilitate evaluation of the effectiveness of pedagogies and instructional designs for improvement, help to monitor students' learning and persistence, predict students' performance, detect undesirable learning behaviours and emotional states, and identify students at risk. Learning analytics can also provide students with insightful information about their learning characteristics and patterns, which can make their learning experiences more personal and engaging, and promote their reflection and improvement.</p> <p>With the accessibility of big data and the digitalisation in the educational setting, Institute of Higher Learning (IHLs) is keen to look at what Learning Analytics may mean to lecturers and students. Particularly, on how teaching and learning experiences may present itself with the adoption of Learning Analytics. Those interested in this project please contact Mr Edwin Seng (edwinseng001@suss.edu.sg) with your CV by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
7	<p>Econometric model building with regression analysis</p> <p>(Supervisor: Dr Tan Khay Boon)</p>	<p>This project allows students to build econometric models using cross sectional data or time series data. Students need to identify the dependent variable of interest and search for explanatory variables to explain the dependent variable, build the econometric model using regression analysis and evaluate the adequacy of the model.</p> <p>Students should be familiar with statistical software such as JMP, SAS or EViews which can perform regression analysis.</p> <p>Those interested in this project please contact Dr Tan Khay Boon at kbtan010@suss.edu.sg with your CV by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
8	<p>Forecasting using time series data</p> <p>(Supervisor: Dr Tan Khay Boon)</p>	<p>This project focus on forecasting the price, quantity, profit, revenue or other quantitative values of a company, an industry or an economy using time series data. Students are expected to obtain the time series data of their interest and build time series model for forecasting the future values. The time series model include regression models, smoothing models and ARIMA models.</p> <p>Students should be familiar with statistical software such as SAS or EViews which can perform regression and forecasts.</p>

		Those interested in this project please contact Dr Tan Khay Boon at kbtan010@suss.edu.sg with your CV by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).
9	Credit Card insights analytics (Supervisor: Zhang Shuai)	<p>Credit card churn analytics is one of the hottest topics in the banking industry. How to leverage current data to predict customer churn trend is critical for any bank. This project will offer students the chance to evaluate a well-prepared Credit Card dataset.</p> <p>Students can step into the project from 3 areas, business case studies, exploratory data analysis (EDA), machine learning techniques (Modelling). In the end, students are expected to provide suggestions to the management team on the business direction based on analytics results.</p> <p>Students who are interested in this project and are conversant with python programming, please contact Mr Zhang at sazhang001@suss.edu.sg with your CV by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
10	Airbnb price & trend analytics (Supervisor: Zhang Shuai)	<p>Airbnb price prediction is very popular across the world especially during pre-pandemic period. To evaluate the industry, providing accurate segmentation and rental prices is quite essential and critical. Students will leverage Airbnb existing dataset online to step into analysis. It's fine for the students to choose any country/city to kick-off the project.</p> <p>Students can step into the project from 3 areas, business case studies, exploratory data analysis (EDA), machine learning techniques (Modelling). In the end, students are expected to provide suggestions to the management team on the business direction based on analytics results.</p> <p>Students who are interested in this project and are conversant with python programming, please contact Mr Zhang at sazhang001@suss.edu.sg with your CV by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
11	Analysis and modelling based on Kaggle datasets (Supervisor: Mr Adam Wong)	<p>This project focuses on data mining or text mining (or the use of both) using Kaggle datasets.</p> <p>Kaggle offers a rich repository of datasets (competition and non-competition) that looks at varying business problem(s). Students have the flexibility to choose which datasets they would like to work on. However, students should note that discussion of business problem is inadequate for most datasets provided on Kaggle, and you will have to develop a convincing business narrative (or non-business for non-commercial datasets) to accompany your analysis and modelling.</p>

		<p>Students can and is free to use any tools for the project. Those interested in this project may contact Mr Adam Wong(adamwong002@suss.edu.sg) by noon 6 Jun 2022 and cc(leeyh@suss.edu.sg).</p>
12	<p>AI predictive modeling of the business expansion of a Singapore logistics company into Southeast Asian countries</p> <p>(Supervisor: Dr How Meng-Leong)</p>	<p>In this project, the student will play the role of a business consultant for a Singapore-based logistics company that is going to expand its business network into other countries in Southeast Asia. There are 3 areas in the business proposal which the company definitely requires for its business expansion plans:</p> <ul style="list-style-type: none"> (1) improvements in logistics business processes (discrete event-based business process modeling), (2) simulation of the logistic flow plan in a typical new warehouse (agent-based modeling), and (3) the selection of 15 new optimal locations for its new warehouses across Southeast Asia (optimization using greenfield analysis). <p>Artificial Intelligence (AI) and 3D simulation make a powerful combination, including when working on training or testing policies and dealing with data quality and quantity. Industrial problems can be resolved by reinforcement learning and simulation. AI can be used on an increasingly wide range of business applications in areas including manufacturing, logistics, supply chains, urban transportation, business processes, healthcare, asset management, and more. Predictive modeling can be utilized to reveal the interplay and tensions between the variables that underlie various business parameters. Computational simulations can be used to produce forecasts of good and bad conditions using multi-variant optimizations. The forecast of these future scenarios is useful for informing policymakers and business stakeholders across domain verticals, so they can make data-driven executive decisions.</p> <p>The free version of Anylogic software may be utilized. Students may also use other software which they are already familiar with. Basic skills in software programming and data analysis of time-series data are required. Knowledge transfer of Artificial Intelligence programming concepts for application in this business analytics project will be provided.</p> <p>Up to 3 students may work on this project, with doing it individually based on (1), (2) and (3). Students who are interested may send their CV to Dr. How (shawnhw001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon 6 Jun 2022.</p>

13	<p>The reliability of Renewable Energy to replace oil and gas as our energy of choice</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>Many renewable energy projects fall short of targets due to weather conditions differing from the forecast or suboptimal performances of the equipment. In 2019 & 2020, 15-20% of the wind and solar projects in India did not meet capacity utilisation targets largely because of wind generation curtailments and lower irradiance for solar projects. From a 2020 Fitch Ratings analysis, it is estimated 90% of wind farms failed to meet their mid-case production levels (P50), likely due to low predictability of wind itself. Wind farms are also estimated to experience 4% of loss in generation from the suboptimal performances of turbines. When not mitigated, the loss from suboptimal equipment performance could be significant: in China, a wind farm commissioned in 2018 only produced 37-45% of its installed potential and a study in revealed the turbine model selection, location of the farm and the turbine hub heights to play a huge part in the farm's losses.</p> <p>Using techniques found in data mining, as well as machine learning and python programming, show how renewable energy production has increased over time, and then determine if it is delivering as expected. Using time series analysis and financial models, determine if the investments made in renewables will pay off over time.</p> <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics, finance and machine learning are preferred. Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
14	<p>Evolution of Greenhouse Gas Production over time</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>A lot of discussion has taken place at the various COP meetings held over the years; the elephant in the room has been and continues to be how the earth will continually warm and the greenhouse gas contribution for developed vs developing countries.</p> <p>However, its also equally important to understand which industries are contributing. The goal of this project is to evaluate the</p> <ul style="list-style-type: none"> (a) 5 highest greenhouse producing countries, (b) 5 lowest greenhouse producing countries, (c) the top 5 industries in each county that contribute to this and (d) predict how the greenhouse gas emission will change over time, based on the evolution of those industries over a 50 year window? <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics and machine learning are preferred. Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>

15	<p>Renewable energy generation and dependency in Singapore</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>Singapore is embarking on a journey to “green up” her sources of energy. In 2021, a team at NTU proposed a study on geothermal energy in Singapore (https://www.ntu.edu.sg/news/detail/study-on-geothermal-energy-potential-for-singapore). However, this is not the only type of renewable energy around. Through bodies like the PUB and the EMA, Singapore has massively “Solarized” our grid (https://www.straitstimes.com/singapore/singapores-first-large-scale-solar-floating-farm-opens-at-tengeh-reservoir). The question becomes – is this enough, to offset our growing demands for energy consumption?</p> <p>Using techniques found in data mining, as well as machine learning and python programming, document (a) Singapore’s energy journey with respect to time, over the past 20 years and then (b) predict, based on trends related to energy consumption and population growth, how much of our power will need to continue to come from non-renewable sources up to 2050.</p> <p>This project will require you to perform a time-series analysis and extract data from https://data.gov.sg or https://www.singstat.gov.sg/ or other such sites to address the question.</p> <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics, energy, environment and machine learning are preferred. You will need to do some financial modelling, so Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
16	<p>Using Data Science to Analyze Solar Power in Singapore</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>Singapore has invested heavily in solar, to offset our carbon emissions and reduce our reliance on fuel imports, so as to meet the energy needs of the country.</p> <p>However, solar by itself is not 100% reliable. Your task in this study is to determine (a) some causes of why solar output is less than expected, (b) estimate the output during over a 1 year window and (c) forecast the output for the next 2 months after.</p> <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics, energy, environment and machine learning are preferred. Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
17	<p>Pressure Prediction for oil and gas drilling and monitoring</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>Pressure management, monitoring and maintenance is one of the most important oil field properties. As oil and gas is produced from the ground, the natural pressure present in the reservoir is reduced and thus different methods of maintaining the pressure have to be devised. To know the optimal point of intervention requires one to monitor the pressure, and the monitoring is often via underground pressure gauges, where readings taken at set points in time. In this way, a “time-lapse” of the pressure depletion is determined.</p>

		<p>However, there are older, legacy wells where it is not possible to install these costly gauges, for a variety of reasons. The wells might not be designed to take these tools, or for reasons of safety and cost, it makes no sense to install new equipment. The only readings in these older wells are in fact those you find on the surface</p> <p>Using methods of machine learning and AI, develop an algorithm where surface measurements can be used to predict the underground pressure value. You will be provided with a data set of publicly available data; you will be required to build an analytical (predictive) model using conventional machine learning algorithms (e.g. decision trees), as well as a model using an artificial neural net (ANN). You will then be asked to perform a comparison and explain which of the 2 models you think is better, and why.</p> <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics, energy and machine learning are preferred. Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
18	<p>Predicting the Rock Type for Mining, Oil and Gas</p> <p>(Supervisor: Dr Munish Kumar)</p>	<p>Oil and gas is typically produced from sedimentary basins which can either be sandstone or carbonates. To produce oil and gas, one must first drill a well, and extract rock samples to determine its storage and flow capacity.</p> <p>Engineers refer to the flow capacity as “permeability” and storage capacity as “porosity”. Traditionally, these 2 terms were used to classify the oil/gas reservoir rock into “rock types”. An oil and gas well may have multiple rock types. In a reservoir model, these rock types are then used to predict away from areas that have been drilled, the goal being to target/drill in locations with the most “optimal” rock type. Unfortunately, just depending on these 2 terms can be problematic as nature is often non-linear. Other important parameters include geographic location as well as geological age of the rock, temperature, pressure, grain density, depth etc. Some geoscientists have thus developed different methods of rock typing, which can be based on relationships to these different terms.</p> <p>Using methods of machine learning and AI, develop classification algorithms to classify rock measurements into unique rock types. You will be provided with data set(s) of publicly available data from different regions around the world; you will be required to build a database (SQL, Tableau, Power BI) which is accessible for use with analytical (predictive) models with at least 2 different clustering methods using machine learning algorithms. You will then be asked to perform a comparison and explain which of the 2 models you think is better, and why.</p> <p>Students who are strong/comfortable with SQL, Python programming and power BI/ Tableau, and interest in data science, analytics, energy and machine learning</p>

		are preferred. Please email Dr. Munish Kumar (munishkumar001@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).
19	Customer Analytics (Supervisor: Ms Tang Yoke Wah)	<p>Building successful customer relationship brings about short-term and, more importantly, long-term profitability to the business. Hence, mature businesses set aside budget and resources for Customer Relationship Management (CRM), which is an integrated approach to identifying, acquiring and retaining customers.</p> <p>The student will conduct techniques applicable for Customer/Marketing Segmentation and generate recommendations for more focused and targeted marketing efforts, across various functions or levels of the organization. Student is to demonstrate deployment of these recommendations with a visualization tool like Tableau, QlikSense, or PBI.</p> <p>Depending on data completeness and availability, student is expected to conduct both supervised (prediction) and unsupervised (dimension reduction and/or clustering) learning techniques with IBM SPSS Modeler or Stats/ JMP/ RStudio/ Python; and map technical outcomes to actionable insights aligned to business goals.</p> <p>Those interested in this project please contact Ms Tang Yoke Wah at ywtang001@suss.edu.sg with your CV by 6 Jun 2022(cc leeyh@suss.edu.sg).</p>
20	The Medical Diagnosis on Dementia in Singapore (Supervisor: Mr Victor Yiew)	<p>There is 1 in 10 above 60 years old has dementia in Singapore. Dementia is not a part of normal ageing. It is a medical condition that causes brain cells to die at a faster rate than normal. The resulting progressive cognitive decline leads to increasing difficulties in coping with everyday activities. A common symptom is poor short-term memory.</p> <p>The symptoms of dementia progress over time and changes across the stages. It may be difficult to place a person with symptoms of dementia separated into three categories: mild, moderate and advanced or in a specific stage as stages may overlap. With increased life expectancy and a rapidly ageing population, the number of people with Neurological diseases such as dementia the fifth leading cause of disability dementia are expected to increase. The importance to raise awareness of dementia, increase support and caregivers is vital in Singapore.</p> <p>The student will use the IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software to uncover medical diagnosis establishing mental wellness and health using https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to translate the formulation on symptom causes and effectual preventive actions to maintain healthy lifestyles.</p> <p>Those interested in this project please contact Mr Victor Yiew at victoryiew002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>

21	<p>Prognostic Detection of Dengue Outbreaks</p> <p>(Supervisor: Mr Victor Yiew)</p>	<p>In 2020, Dengue death toll hits a record high of 28 fatalities amid the worst-ever outbreak... The number surpasses the previous high 15 years ago, when full-year dengue fatalities hit 25, with the dengue death tally now matching the number of Covid-19 fatalities here. Singapore is still in the midst of the yearly traditional dengue peak season, which usually takes place between May and October, when weather is warmer. What could be done to prevent more death?</p> <p>In 2013, Singapore reported largest dengue fever outbreak in history recording some 22,318 cases and eight deaths. However, that record was easily shattered in 2020 as Singapore officials report that as of Dec. 17, 2020, 34,844 dengue cases were reported. The factors included the prevalence of a less common dengue strain - DENV-3 - female Aedes aegypti mosquitoes are day-biters and harbor within indoor environment. The work-from-home arrangement to allow social distancing has led to most people working from ventilated homes. This increased the probability of contact between Aedes aegypti mosquitoes and humans and would have contributed to the increase in dengue cases.</p> <p>The student will use the IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software to exploit the reduce of dengue cases from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to develop on insightful Infographics and through Story Showing on socio-demographic, environmental factors, health status and life habits, and government schemes and support to reduce the number of dengue death in Singapore.</p> <p>Those interested in this project please contact Mr Victor Yiew at victoryyiew002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
22	<p>Envigor Circular Economy to City of Our Future.</p> <p>(Supervisor: Mr Victor Yiew)</p>	<p>Singapore is persistently and geologically challenged and with climate change and growing population and economic competition - the country needs land for urban development. To overcome the pervasive space challenge, the country is investing in megaprojects that will help overcome these problems and take us into the next century.</p> <p>Tomorrow city is near, construction of Deep Tunnel Sewerage System on such a scale comes with challenges to build within a restricted space employs state of the art machinery and technology; renewable energies such as solar power part of Singapore's solar masterplan to meet its targeted energy output by 2030 have relook on how land space effectual utilize; our city of tomorrow is working hard to secure its economic future by building multi-billion dollar Tuas megaport to support multiplex projects - with consideration of circular economy approach.</p> <p>The student will use the IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework to determine opportunity of Upcycle recycle of quantifiable effects on physical substance been used or produced or consolidating of data from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to classify</p>

		<p>and/or develop analytical techniques to protect the earth and secure available resources for land development.</p> <p>Those interested in this project please contact Mr Victor Yiew at victoryyiew002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
23	<p>The Prophecy of Data Scientist Remunerations</p> <p>(Supervisor: Mr Victor Yiew)</p>	<p>How Much is a Data Scientist Worth in 2020? In these unprecedented times of uncertainty, current events have shown just how valuable the fields of Data Science and Computer Science truly are - dashboard, contact tracing, and data analytics - compose the “virtual front lines” of our attack on the pandemic and continuously prove to be driving sources of change. However, one question still remains: Exactly how much are Data Scientist valuable for?</p> <p>What is it Like to Be a Data Scientist in 2021? Rise in demand for data scientist continues in 2021 too - From established companies like Accenture, Standard Chartered Bank, to social media giants Facebook and Agoda – all the way up to hospitality and hospital management – companies are hiring data scientists! But what is it that makes this role the “Sexiest Job Role of the 21st century”?</p> <p>The student will use the IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework to predict the remuneration package of data scientist and/or data analyst benefit security from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to accolade the useful sources of information to showcase integral Data Scientist financial career roadmap.</p> <p>Those interested in this project please contact Mr Victor Yiew at victoryyiew002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
24	<p>The Credit Card Fraud Detection</p> <p>(Supervisor: Mr Victor Yiew)</p>	<p>Credit card frauds are more common with scammers online, messages and call on direct contact through mobiles, and lately, they’ve been on the higher side. To put this in simple terms, these frauds path will cross a billion credit card users by the end of 2022. The innovations in technologies like Artificial Intelligence, Machine Learning, and Data Science, credit card companies have been able to successfully identify and intercept these frauds with sufficient accuracy.</p> <p>The idea here is to analyze the customer’s usual spending behavior, including mapping the location of spending to identify the fraudulent transactions from the non-fraudulent ones. For this project, make use of customer’s transaction history as dataset and ingest it into decision trees, Artificial Neural Networks, and Logistic Regression. As you feed more data to your system, you should be able to increase its overall accuracy, so do the abilities to utilize it in different ways, including when</p>

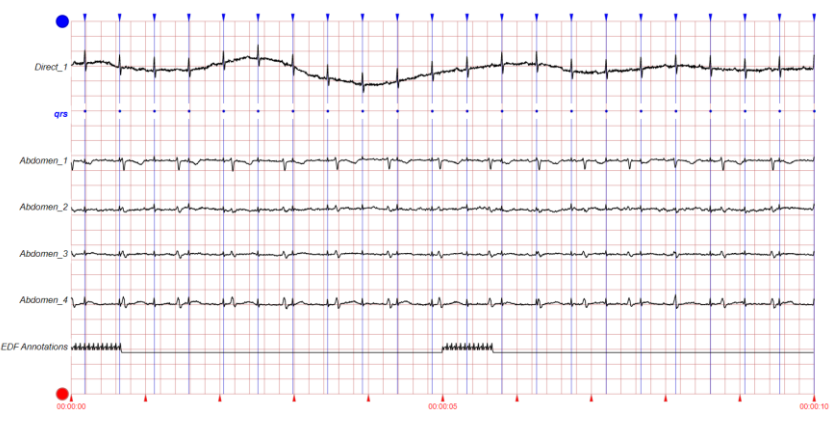
		<p>making predictions for the future, otherwise known as prescriptive analytics of finance.</p> <p>The student will use the IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework to examine dataset from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to derive on insightful and actionable value proposition.</p> <p>Those interested in this project please contact Mr Victor Yiew at victoryiew002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
25	<p>Classification Model of Housing Price Index Movement Based on Past Market Data & Economic Indicators</p> <p>(Supervisor: Dr Teh Yong Liang)</p>	<p>This project aims to build a classification model to predict the future movement of a housing price index based on past market data such as pipeline supply of housing, vacancy rate, rental price growth, population growth and household income growth etc.</p> <p>The student will be tasked to build and study various models for the classification of the index movement, such as by CART, logistic regression, random forest as well as ensemble models. A time-invariant approach should also be investigated where derived predictors involving changes will be considered instead of absolute values. The output of the model would be classes ranging from strong growth to severe decline of the housing price index. Such a model would be useful to assist policy makers in the enactment of cooling or loosening measures for the stability of the housing market.</p> <p>(The IBM SPSS modeler or other suitable software can be used for this project.)</p> <p>Those interested in this project please contact Dr Teh Yong Liang at ylteh002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>

26	<p>Predictive Modelling of Market Index Direction Based on Past Economic Indicators</p> <p>(Supervisor: Dr Teh Yong Liang)</p>	<p>This project aims to build a classification model to predict the future direction of a market index, such as the S&P 500, based on past economic indicators such as GDP growth rate, unemployment rate, inflation, interest rate and consumer confidence etc.</p> <p>The student will be tasked to investigate several variations of the model to identify the 'best' model. One variation would be using past years' indicators to predict the present year's index movement. Another would be to use past quarters' indicators to predict the present quarter's index movement. Or it could also be a combination of indicators over different time periods as well as including derived indicators. The output of the model could be several classes, ranging from strong downturn to strong rally. Such a model could potentially help investors to avoid large losses and/or increase gains.</p> <p>(The IBM SPSS modeler or other suitable software can be used for this project.)</p> <p>Those interested in this project please contact Dr Teh Yong Liang at ylteh002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
27	<p>Identification of Illegal Smoking by ANN Recognition Model</p> <p>(Supervisor: Dr Teh Yong Liang)</p>	<p>Despite many years of public education and enforcement by NEA, illegal smoking remains prevalent in Singapore which results in health hazards caused by inhalation of second-hand smoke.</p> <p>This project aims to build an ANN model capable of recognising smoking activity from camera surveillance video or images.</p> <p>The student will be tasked to research image recognition models, such as convolutional neural networks, and use them to build a classification model capable of smoking detection. It is also expected that a test deployment of the trained model be applied to a webcam to demonstrate a live detection of smoking activity. If successful, such a model can be applied to camera surveillance in various parts of Singapore to assist the authorities in the enforcement of smoke-free environments.</p> <p>(Python or R programming can be used for this project.)</p> <p>Those interested in this project please contact Dr Teh Yong Liang at ylteh002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>

28	<p>Short Term Trading Bot of Security by Reinforcement Learning</p> <p>(Supervisor: Dr Teh Yong Liang)</p>	<p>This project aims to build a short term trading bot automatically executing day trades purely based on price indications of the underlying security as well as derivative indicators such as options or futures pricing. Other technical indicators and/or a time-invariant approach can also be considered. Because of the stochastic nature of stock prices over short time frames, a reinforcement learning model might be more viable over other predictive models based on supervised learning.</p> <p>The student will be tasked to research on reinforcement learning concepts and models, such as Q-learning and A3C models. The reinforcement learning model would be based on an agent trained on using current state variables to execute within an action space of either a buy, sell or do nothing actions, with positive or negative rewards to the agent dependent on the future price action of the security. After the bot is trained, it should be backtested on historical data over several time-periods to assess its robustness and performance.</p> <p>(Students should note that this project is more challenging. Python or R programming can be used for this project.)</p> <p>Those interested in this project please contact Dr Teh Yong Liang at ylteh002@suss.edu.sg with your CV by 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>
29	<p>Sales Analytics for Engineering Services</p> <p>(Supervisor: Dr Ren Jing)</p>	<p>This project requires the student to work with Quest Technology. Quest Technology is a regional leader in micro-contamination control engineering services, with branches in many countries and regions in Asia, and its customer base includes SIEMENS, Hyflux, DSO, HP, Hitachi, etc. The main goal of this applied project is to help Quest Technology to better understand and facilitate their marketing and sales.</p> <p>Students will take initiative to visit the company and discuss/work with the professionals on the project. They are expected to help Quest Technology to collate marketing and sales relevant data (from their database), mine sales patterns, and/or propose useful deployment suggestions to help the company increase their sales in a region or sales of a specific product. This is a good opportunity for students to use real data and become familiar with how theory (e.g., visualisation, clustering, or predictive modelling) can be used in practice.</p> <p>This project will be co-supervised by Dr Ren Jing and Dr Carmen Lee Kar Hang. Students who are keen on exploring valuable patterns from raw data are to send their CVs to them (jingren@suss.edu.sg, carmenleekh@suss.edu.sg) and copy leeyh@suss.edu.sg by 6 Jun 2022. Students who are working in the same industry as this company will not be selected for this project to avoid any conflict of interest.</p>

30	<p>Predictive analytics for industrial properties</p> <p>(Supervisor: Dr Carmen Lee)</p>	<p>This project involves working with SN Real Estate, a real estate company dealing with the sales of commercial and industrial properties. The company is interested in creating a predictive model to estimate prices of industrial properties using potential drivers that include location, usage type (B1-light industry; B2-heavy industry), structure type (ramp up, flatted or landed), ceiling height, size, year etc.</p> <p>This project involves analysing the data hosted by the company with the use of predictive modelling. Students are expected to conduct a literature review on (i) potential factors affecting industrial property prices and (ii) possible machine learning algorithms for predictive modelling in the given context. Then, to remove bias, multiple algorithms should be employed to build the predictive models. The project deliverables are expected to provide a better understanding of the sales market for the company.</p> <p>Those interested please send CV to Dr Carmen Lee at carmenleekh@suss.edu.sg and cc A/P Lee Yew Haur at leeyh@suss.edu.sg by noon 6 Jun 2022. Students who are working in the same industry as this company may not be selected for this project to avoid any conflict of interest.</p>
31	<p>Time Series Forecasting for Stock Indices or Forex Rates</p> <p>(Supervisor: Yeo May Peng)</p>	<p>Perform modelling and forecasting of time series using the ARIMA/GARCH models.</p> <p>Option 1: Predict price movements of stock indices such as the Straits Times Industrial Index and correlate with the underlying basket of stocks.</p> <p>Option 2: Predict foreign currency exchange rate movements and correlate with the economic indicators.</p> <p>Option 3: Forecast financial time series of your choice approved by the supervisor.</p> <p>Student will use Python and Jupyter Notebook to break down the trend, seasonal and residual components. Check stationarity in time series using statistical methods such as mean, variance and the ADF test. Apply ARIMA forecasting model and GARCH model for volatility with smoothing techniques to remove noise. Evaluate forecast results from the model and propose ways to make improvements.</p> <p>Interested student who are proficient in Python may contact Yeo May Peng at mpyeo002@suss.edu.sg with your CV and cc A/P Lee Yew Haur at leeyh@suss.edu.sg by noon 6 Jun 2022.</p>
32	<p>Crypto and Stock Analytics</p> <p>(Supervisor: Mr Oh Chin Lock)</p>	<p>Cryptocurrencies have experienced multiple short-term rallies and crashes in recent years. Are cryptocurrencies really assets that investors should seriously look at vis-a-vis equities/stocks, or are they like “a random walk down wall street”?</p> <p>This project will study the use of various sequential modelling/time series techniques to try to model the movement of cryptocurrencies and compare them against traditional stocks and trading strategies. The main objective is to understand any correlations between stocks and cryptocurrency and try to model</p>

		<p>their movements using various techniques and recommend a plausible portfolio/trading strateg (if any). Techniques that could be explored include ARIMA, deep neural network techniques, classification, as well as genetic algorithms. Students need to know how to program in python to work on this project.</p> <p>The main objective is not to predict future crypto or stock prices but to gain a better grasp of various modelling techniques that can be used. As a secondary objective, the student will also self-study to gain more in-depth knowledge of fintech like crypto, blockchain, tokens, and Defi.</p> <p>Those interested should submit their CVs to Mr Oh(cloh001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon on 6 Jun 2022.</p>
33	<p>Summarizing Responsible AI Landscape using NLP</p> <p>(Supervisor: Mr Oh Chin Lock)</p>	<p>The rapid advances in AI and adoption in many areas have led to concerns among the public and users about it's proper and responsible use. In that light, there are many on going governance frameworks being developed around the world to try to regulate and mitigate the risks in AI deployment. The focus is in putting the human back in charge so that AI's use can be trusted.</p> <p>This project will use NLP text-mining and visualization techniques to try to summarize the state of the art in the Responsible AI landscape. Students can make use of datasets that include research databases, patent databases, job databases, governance publications, as well as news articles and tweets of influential AI thought leaders to understand various aspects of the landscape. Useful reference: https://www.weforum.org/projects/ai-ethics-framework</p> <p>After completing the project, students should not only develop a strong understanding of NLP techniques but also obtain a firm grasp on Ethical and Responsible AI principles and developments around the world. Students need to know how to program in python to work on this project.</p> <p>Those interested should submit their CVs to Mr Oh(cloh001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon on 6 Jun 2022.</p>
34	<p>Baby's Fetal ECG during pregnancy</p> <p>(Supervisor: Mr Oh Chin Lock)</p>	<p>There are currently a few ways a gyneacologist tracks the development of a baby throughout a mother's pregnancy. One of the ways is to use non invasive fetal ECG (electro cardiogram) via electrodes placed on the mother's abdomen. However, such ECG signals are usually complex, containing not only weak fetal ECG signals (which is what the doctors want to see) but also the mother's ECG</p>

		<p>signals as well as other ‘noise’ such as mother’s respiration, baby’s fetal movements, etc. An ECG signal is like a ‘time series’.</p>  <p>The objective of this project is to try to develop machine learning techniques to extract weak fetal ECG signals from complex ECG signals measured from the mother’s abdomen. Techniques could include filtering, PCA, advanced neural network techniques or others. Students need to know how to program in python to work on this project.</p> <p>Those interested should submit their CVs to Mr Oh(cloh001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon on 6 Jun 2022</p>
35	<p>Modelling, analysis and regression based on time-series data</p> <p>(Supervisor: Dr Chris Ho)</p>	<p>This project focuses on the analysis of time-series data such as revenue, market and stock pricing using ARIMA models, and prediction of future trends using time-series regression.</p> <p>Students who are strong/comfortable with Python/R programming, and interest in data science, analytics and machine learning are preferred. Please email Dr. Chris Ho (jhho003@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>
36	<p>Product recommendation s based on sentiment analysis/natural language processing</p> <p>(Supervisor: Dr. Chris Ho)</p>	<p>This project focuses on developing a recommender system based on product reviews, leveraging on natural language processing and sentiment analysis.</p> <p>Students who are strong in Python/SPSS modeler, and interested in data science, analytics, and machine learning are preferred. Please email Dr. Chris Ho (jhho00@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg).</p>

37	<p>Discovery of key business/economic trends from social media using natural language processing</p> <p>(Supervisor: Dr. Chris Ho)</p>	<p>This project focuses on the analysis of social media feeds (e.g. Facebook, Twitter, Youtube, TikTok, Instagram, Linkedin, Reddit) to discover key business/economic trends, leveraging on natural language processing.</p> <p>Students who are strong in Python/SPSS modeler, and interested in data science, analytics and machine learning are preferred. Please email Dr. Chris Ho (jhho003@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 and cc (leeyh@suss.edu.sg)</p>
38	<p>Text Mining on Qualitative Survey Responses</p> <p>(Supervisor: A/P Lee Yew Haur)</p>	<p>Two students will work with SUSS to use text mining on the qualitative responses to the Graduate Employment Survey 2021. Each student will tackle 1 qualitative question.</p> <p>The students will use IBM SPSS Modeller to categorise the qualitative response into meaningful and actionable categories and also perform further analyses with structured data to uncover further insights. The students can further extend the analyses using R, if they have time.</p> <p>Those interested should submit their CVs to A/P Lee Yew Haur (leeyh@suss.edu.sg) by noon on 6 Jun 2022. Selection will be competitive.</p>
39	<p>Text Mining on Qualitative Survey Responses</p> <p>(Supervisor: Mr Adam Wong)</p>	<p>Two students will work with SUSS to use text mining on the qualitative responses to the Graduate Employment Survey 2021. Each student will tackle 1 qualitative question.</p> <p>The students will use IBM SPSS Modeller to categorise the qualitative response into meaningful and actionable categories and also perform further analyses with structured data to uncover further insights. The students can further extend the analyses using R, if they have time.</p> <p>Those interested should submit their CVs to Mr Adam Wong(adamwong002@suss.edu.sg) and cc (leeyh@suss.edu.sg) by noon on 6 Jun 2022. Selection will be competitive.</p>
40	<p>Learning Analytics Project</p> <p>(Supervisor: Li Jizhi)</p>	<p>Supporting student learning has always been a critical part of higher education, which plays a pivotal role in helping the students towards achieving greater academic success. The rich pool of student data allows the educational institutions to mine useful insights for various purposes. Support for student can be provided when they are newly enrolled, to help them adapt to the new environment. Educators can provide academic support to those who may potentially underperform and catch them before falling. Interventions can also be targeted at graduating students, to uncover the determinants of successful graduation and nudge students to excel further.</p> <p>Student can choose a more specific subtopic to dive into, translate data into actionable insights that educational institutions can leverage on to better support their students.</p>

		<p>Students should be familiar and comfortable with analytics tools such as IBM SPSS Modeler/R/Python.</p> <p>Interested students may contact Ms Li at jzli002@suss.edu.sg with your CV by 6 Jun 2022(cc leeyh@suss.edu.sg).</p>
41	<p>Analysing and Forecasting COVID-19 Case Loads and Fatalities</p> <p>(Supervisor: Dr Goh Shao Hung)</p>	<p>This project will focus on analysing the COVID-19 global dataset from the past two years, with the objectives of obtaining insights from the past time-series data and generating forecasts for the future.</p> <p>While COVID-19 is a highly unpredictable disease, some insights could already be discernible from the past two years. For example, COVID-19 caseloads have shown strong periodicity of about 4 months, due to countries' cycles of lockdowns and loosening, as well as the emergence of new variants and the subsequent development of natural immunity in a population. Furthermore, deaths from COVID-19 tend to lag caseloads by 2 to 3 weeks.</p> <p>A forecast of COVID-19 caseloads and fatalities will be useful in many contexts, for example to derive forecasts for the demand of hospital beds, vaccines, surgical gloves, face masks and hand sanitisers. The development of a forecasting model to predict the outcomes of future pandemics will also be of great interest to governments and international healthcare organisations.</p> <p>The potential project scope may include the back-testing of forecasting algorithms, tuning of model parameters, exploration of noise reduction techniques such as low-pass filters and investigation of causal factors in a forecast model. Suitable tools for analysis would include Python, R, SPSS and/or other data visualisation software.</p> <p>Past two years' data for the project is publicly available at https://ourworldindata.org/coronavirus</p> <p>Please email Dr Goh Shao Hung (shgoh015@suss.edu.sg) directly to express your interest in working on this project by noon 6 Jun 2022 (cc leeyh@suss.edu.sg).</p>

Updated on 2 Jun 2022