

ANL488 Project List for Jul 2023 Semester

Those interested in any of the projects, please email the respective supervisors with your CV by noon 5 Jun 2023 and cc'ed jesstanwc@suss.edu.sg.

No	Project Title	Supervisor	Description
1	Sentiment Analysis on product/services at Digital Marketplace	Mr Chua Poh Chai pcchua002@suss.edu.sg	<p>Problem Statement: In the digital marketplace, product/service reviews are captured extensively. However, other than the “number of stars” which is quantitative and measurable, most of the feedback are in text. The vast amount of unstructured data (e.g. text) are not gainfully and actively analysed and used. Sentiment Analysis processes the text and sieves out the relevant parts of texts–sentiment on a product feature or functionalities of a product, thereby helping decision makers with insights to improve their product/services offerings. For more, refer to https://monkeylearn.com/sentiment-analysis/section-on-Market-Research.</p> <p>This project aims to process the text from product/service review into intelligence and develop sentiment rating (score).For an example, see https://monkeylearn.com/blog/sentiment-analysis-of-product-reviews/Students are expected to develop their algorithms and codes. With today market’s immense demand for Natural Language Processing (which Sentiment Analysis is a subset) talents , students with the skills and experience in this area will stand in good stead.</p> <p>Students who are keen in this project can email Chua Poh Chai at pcchua002@suss.edu.sg with their CV and copy jesstanwc@suss.edu.sg. As this project focuses on sentiment analytics, students’ knowledge in statistics, text-mining and machine learning, together with strong Python/R programming, will be needed</p>
2	Anomaly Detection in Time Series Data	Mr Chua Poh Chai pcchua002@suss.edu.sg	<p>Anomaly detection is one of the most critical functions for companies and decision makers who monitor their portfolio and receive alerts due to abnormal observations in the data. These include system spikes or failures in IT resource management, changes (sudden or over time) in e-commerce traffic, fraud detection in banking transactions, financial market data on economy/industry/company and health tracking devices (health monitoring devices in smart watches or others) and others.</p> <p>This project aims to create anomaly detection models using machine learning (ANN) and statistical analysis. Both supervised and unsupervised methods will be used. Students are expected to collect quantitative time series data and develop their algorithms and codes.</p> <p>Students who are keen in this project can email Chua Poh Chai at pcchua002@suss.edu.sg with their CV and copy jesstanwc@suss.edu.sg. As this project focuses on statistical analysis, students’ knowledge in statistics and machine learning, together with strong Python/R programming, will be needed.</p>

3	Predictive analytics for industrial properties	Dr Carmen Lee carmenleekh@suss.edu.sg	<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. Most importantly, the project output should enable the company to answer the question: what will be the price of a particular property this year?</p> <p>Regular meetings with the company's representatives are expected. Those interested please send CV to Dr Carmen Lee at carmenleekh@suss.edu.sg. In your CV, please provide the grades you obtained in ANL courses. 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>
4	Predictive analytics for food safety violations	Dr Carmen Lee carmenleekh@suss.edu.sg	<p>This project is a meaningful project that can help achieve greater public health. Many foodborne illness outbreaks originate in restaurants. To prevent these outbreaks, health authorities in many countries have been conducting restaurant inspections regularly. However, the effectiveness of the current inspection policy remains an open question as we often witness foodborne illness outbreaks. On the other hand, customers who feel sick after visiting a restaurant may not report to health authorities but disclose this by posting an online review on social media. Therefore, by monitoring social media discourse, health authorities might get valuable information to identify high-risk restaurants for timely inspection.</p> <p>This project is to leverage customer reviews to predict food safety violations of restaurants. Students are expected to look for open-source data related to restaurant inspection records (e.g., from NYC Open Data) and also do web-scraping to collect relevant reviews from social media platforms such as Yelp and TripAdvisor. Text mining is required to extract some textual features from the reviews, and use them as inputs of predictive models to predict if a restaurant violates the regulations.</p> <p>Students who are good at Python programming, have knowledge/experience in web-scraping, data transformation, labelling, procedural programming, Pandas and Jupyter, are needed. Those interested please send CV to Dr Carmen Lee at carmenleekh@suss.edu.sg. In your CV, please provide the grades you obtained in ANL courses.</p>
5	Modelling, analysis and regression based on time-series data	Dr Chris Ho jhho003@suss.edu.sg	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. Students who are strong/comfortable with Python/R programming, and interest in data science, analytics and machine learning are preferred.
6	Product recommendations based on sentiment analysis/natural language processing	Dr Chris Ho jhho003@suss.edu.sg	This project focuses on developing a recommender system based on product reviews, leveraging on natural language processing and sentiment analysis. Students who are strong in Python/SPSS modeler, and interested in data science, analytics, and machine learning are preferred.

7	Discovery of key business/economic trends from social media using natural language processing	Dr Chris Ho jhho003@suss.edu.sg	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. Students who are strong in Python/SPSS modeler, and interested in data science, analytics, and machine learning are preferred.
8	Logistics Performance Index 2022	Dr Goh Shao Hung shgoh015@suss.edu.sg	<p>The Logistics Performance Index (LPI) is an international benchmark that measures countries on their logistics efficiencies. Developed by the World Bank, the 2022 edition of the LPI is expected to be published by late 2022 or early 2023. The objectives of this project would be 1) to explore trends in the latest LPI dataset against results from previous editions of the LPI and 2) to investigate the links between a country's social-economic parameters and its logistics performance over time. Relevant social-economic parameters might include institutional quality, government spending on infrastructure and depth of human capital (all of which are publicly available).</p> <p>Possible methodologies for the project may include 1) regression or cluster analysis using SPSS Modeler and other similar tools; 2) the development of a classification or predictive model for LPI performance using Python or R; and/or 3) the development a dashboard that presents the past/current LPI and related datasets in an interactive way using PowerBI or Tableau.</p> <p>The results of such an analysis can aid in the uncovering the source of a country's competitiveness over time and versus other countries. It also has implications on businesses' decision-making on international trade and allocation of investments. Depending on your strengths and interests, the scope may be divided and the project be attempted more than one student.</p>
9	3D computational modeling of the business expansion of a Singapore logistics company into Southeast Asian countries	Dr How Meng-Leong shawnow001@suss.edu.sg	<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> <p>(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 10 new optimal locations for its new warehouses across Southeast Asia (optimization using greenfield analysis).</p> <p>Computational modeling 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. These techniques 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 is required.</p> <p>Students may work on this project, doing it individually based on (1), (2) and (3).</p>

10	Modelling and Forecasting Time Series	Dr Karl Wu karlwuky@suss.edu.sg	<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. We will be using either SAS Forecast Studio and/or R for this project.</p>
11	Forecasting of industrial property prices	Dr Karl Wu karlwuky@suss.edu.sg	<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 forecasting the prices of different types of industrial properties using historical data.</p> <p>This project involves modelling time series data of the property prices in the past 20 years. Students should have completed ANL317 Business Forecasting and a good understanding of concepts such as autocorrelation, exponential smoothing, ARIMA, etc. SAS or R can be used for analysis and data visualisation.</p> <p>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>
12	The reliability of Renewable Energy to replace oil and gas as our energy of choice	Dr Munish Kumar munishkumar001@suss.edu.sg	<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.</p>

13	Renewable energy generation and dependency in Singapore	Dr Munish Kumar munishkumar001@suss.edu.sg	<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.</p>
14	The Interplay between Investor Expectation, Governance and Regulations in Carbon Markets and Pricing	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>We know for a fact that investors are demanding greater accountability from governments and companies to minimize their carbon footprint. In this case, what often happens is investors influence markets by divesting from companies that are viewed to have poor carbon management, while governments act to penalize companies by enacting various means of carbon pricing and taxation. The logic here is that both parties act to squeezing companies from both ends to behave and conform to market expectations</p> <p>Companies of course don’t seek to lose investment capital, so they enact methods to reduce their carbon footprint e.g. by voluntarily increasing investment in “greening projects”. In this way, the hope is to offset their carbon footprint, and maybe generate some excess carbon credits, which they can then sell to other companies, thereby turning a problem into a solution. The flip side is that some companies are unable to do so, in which case the government steps in to enforce penalties and taxes, as a signal to improve green practices.</p> <p>Using techniques found in data mining, as well as machine learning and python programming, perform web scaping of a series of research, news and website articles to (a) determine the most effective companies/countries that are able to execute this, (b) companies/countries that are the least effective? (c) If applied to the Singapore business environment, what are the key learnings/lessons that we can take away?</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.</p>

15	Use of ANN to predict Recovery Factor For Oil and gas Production	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>The estimation of recoverable hydrocarbons, or field recovery factor (RF), is a critical variable for Oil and Gas (O&G) companies to identify profitable investments, plan and optimise field development, manage/ monitor ongoing production, and rank commercial and technical decision making. However, determination of RF remains one of the greatest uncertainties in O&G projects, particularly in the early stages of field life. Different methods have been attempted to try and predict field RF, from things as varied as compiling analogs to others as complex as machine learning and artificial intelligence</p> <p>Using techniques found in artificial intelligence and machine learning, students are required to build and compare a series of neural network models, finding the best optimised solution to the question of how one goes about predicting RF based on a series of inputs. Students will be provided with 2 data sets, and will also be required to perform a comparison across both data sets with simpler regression based methods.</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.</p>
16	Do we perceive oil and gas, renewables and our energy future differently if our views are expressed in a language other than English	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>Our views are often dominated by our exposure to information. Information can come in many forms, whether its images, numbers or text, but the context is formed by our own experiences and knowledge. What this means is that information must be in a form we can understand, breakdown and digest, and compartmentalise in our psyche. Take for example individuals who are predominantly "native" English speakers. Such individuals will have their views and opinions formed based on information they receive in a language they understand the best, being in this case English.</p> <p>Take the opinions expressed by investors and politicians regarding oil and gas companies. Western media often portrays companies like Shell and BP negatively, and as ardent consumers of western based media, we sometimes share the same view, partly because such media (be it social or traditional forms) informs us that they are so. However, would this be the same in countries like China, France, Spain, Indonesia etc? Do the Chinese hold this view of CNOOC? How about the French with Total? And how does this link to the discourse ongoing with renewables.</p> <p>In this work, students will use techniques of multilingual sentiment analysis to form a comparative view of how English vs non-English speakers view oil companies, global warming, the energy transition and renewables. Students must web scrape information sites that are not in the native English as well as English language sites. They can use any form of media they wish (twitter, reddit, weibo etc) but the goal is to determine what differences exist, if any.</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. As the student will be using non-English language as well, it would be good if the student is bi- or even tri-lingual.</p>

17	Prediction of potential CO2 storage locations	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>There is great drive and impetus to propose and develop carbon capture sites, seen as a near term viable solution to capture atmospheric and produced CO2. The drive is coming both from government and investors interested in addressing this challenge of global warming.</p> <p>However, these projects are typically large and complex and require quite a bit of planning beforehand. Additionally, the space is constantly evolving in terms of deployable technologies, evaluation of sites, estimation of project costs and just the amount of CO2 that can be captured and held. Therefore, there is a huge push now to share as much public resource as possible to encourage companies to assume the risk of developing carbon capture projects. Some examples of public links are: https://www.ogci.com/co2-storage-resource-catalogue/co2-data-download/ https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database https://setis.ec.europa.eu/european-co2-storage-database_en</p> <p>Using publicly scrapped data, students must develop a machine learning model which would allow one to predict viability of a site for carbon capture using information put together from these public databases. This model should be able to predict the outcome in terms of estimated project costs and the amount of CO2 it can potentially hold.</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.</p>
18	Using machine Learning to uncover insights related to safety incidents	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>Safety is one of the most important things within any large organisation. We all want to work to earn a living, but the important thing is to be able to come back home safely to our loved ones. Safety incidents are especially prevalent in large engineering companies, be it in construction, mining or oil and gas. The study you will undertake is to see how you can improve the safety for employees working within a large mining organisation.</p> <p>You will be provided with datasets relating to the historical safety incidents themselves, site production, employee rosters / leave / duration, and much much more! Can you use this data to help improve the safety for all current and future employees?</p> <p>The student will be tasked with uncover insights that reveal useful information around the safety of mine-site employees / contractors. Most importantly, these insights must be <u>actionable i.e. useful information within your analysis that can practically be utilised to improve the safety performance across the business.</u></p> <p>Students who are strong/comfortable with Python programming and power BI/ Tableau, and interest in data science, analytics, energy, environment, health and safety and machine learning are preferred.</p>

19	Optimisation of business portfolio	Dr Munish Kumar munishkumar001@suss.edu.sg	<p>Oil and gas Companies often look to right size their business portfolio to make sure that they have a well balanced mix of oil and gas assets, and also do not have assets that bring in very low amounts of revenue and cost a lot to run. It is often the strategy teams job to determining which assets to put up for sale/divest, but teams are often faced with the challenge of having not enough time, or too many assets of differing type, and suffer from information overload, resulting in poor decision making. Additionally, they are sometimes required to perform benchmarking, but this is a very manual and tedious process, and thus, a computational solution would be very much preferred.</p> <p>For this project, students will be provided with a data set from various large oil and gas companies, with liquid and gas production information and costs. Students will be tasked with optimising the portfolio based on cost and production yield/mix. The student will also be asked to benchmark across companies</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.</p>
20	Econometric model building with regression analysis	Dr Tan Khay Boon kbtan010@suss.edu.sg	<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>
21	Forecasting using time series data	Dr Tan Khay Boon kbtan010@suss.edu.sg	<p>This project focuses 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. Students should be familiar with statistical software such as SAS or EViews which can perform regression and forecasts.</p>
22	Classification Model of Housing Price Index Movement Based on Past Market Data & Economic Indicators	Dr Teh Yong Liang ylteh002@suss.edu.sg	<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. The model can also consider the possible leading effect of the inputs to the output, where the inputs can be staggered (in time) to the output to order to build a 'predictive' classifier.</p> <p>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, and also help investors in forecasting the property market.</p> <p>(The IBM SPSS modeler or other suitable software can be used for this project.)</p>

23	Predictive Modelling of Market Index Direction Based on Past Economic Indicators	Dr Teh Yong Liang ylteh002@suss.edu.sg	<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. A time-invariant model that only takes in change inputs can also be investigated. The output of the model could be several classes, ranging from strong downturn to strong rally.</p> <p>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>
24	Identification of Illegal Smoking by ANN Recognition Model	Dr Teh Yong Liang ylteh002@suss.edu.sg	<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. With the recent popularity of Chatgpt, the transformer network can also be investigated in such a image-recognition task. 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>
25	Short Term Trading Bot of Security by Reinforcement Learning	Dr Teh Yong Liang ylteh002@suss.edu.sg	<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 hold 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>

26	Predictive Modelling of Asset Prices by Transformer Networks in Generative AI	Dr Teh Yong Liang ylteh002@suss.edu.sg	<p>This project aims to build a predictive model primarily based on the transformer network architecture which powers ChatGPT. The input will be a sequence of asset prices as a time-series and the output will be the asset price in the next time frame. Transformer networks can be suitable in this context because asset prices and language processes are similar in the sense that the transformer takes in a sequence (of prices or words) and outputs the next (price or word).</p> <p>The student will be tasked to research on transformer networks, especially the self-attention mechanism which is central to the performance of the transformer. Then, a predictive model will be built based on price-sequence inputs and trained using an unsupervised (or self-supervised) approach to output the next price. It is also possible to concatenate each price together with other indicators such as traded volume or moving averages etc to create a vector input sequence. The number of cascaded transformers in creating the best performing model can also be investigated, along with other hyper-parameters such as the input sequence length etc.</p> <p>(Python or R programming can be used for this project.)</p>
27	Predictive Modelling of FX Swap Rates by Transformer Networks in Generative AI	Dr Teh Yong Liang ylteh002@suss.edu.sg	<p>This project aims to build a predictive model primarily based on the transformer network architecture which powers ChatGPT. The input will be a sequence of FX swap rates as a time-series and the output will be the swap rate in the next time frame. Transformer networks can be suitable in this context because swap rates and language processes are similar in the sense that the transformer takes in a sequence (of rates or words) and outputs the next (rate or word).</p> <p>The student will be tasked to research on transformer networks, especially the self-attention mechanism which is central to the performance of the transformer. Then, a predictive model will be built based on sequence of rate inputs and trained using an unsupervised (or self-supervised) approach to output the next rate. It is also possible to concatenate each rate together with other indicators such as balance of payments and GDPs etc to create a vector input sequence. Considering the connectedness of currency exchanges, another approach would be to input a vector of various currency pair rates and output the corresponding vector for the next time frame.</p> <p>The number of cascaded transformers in creating the best performing model can also be investigated, along with other hyper-parameters such as the input sequence length etc.</p> <p>(Python or R programming can be used for this project.)</p>
28	Analysis and modelling based on Kaggle (or any publicly open source) datasets	Mr Adam Wong adamwong002@suss.edu.sg	<p>This project focuses on data mining or text mining (or the use of both) using Kaggle (or any publicly open source) datasets. 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 (or any publicly open source), and you will have to develop a convincing business narrative (or non-business for non-commercial datasets) to accompany your analysis and modelling. Students can and is free to use any tools for the project.</p>

29	Advanced analytics to assess Corporate Risk and predict default	Mr Chua Poh Chai pcchua002@suss.edu.sg	<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>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>
30	Learning Analytics: Predictive Model on at-risk students	Mr Edwin Seng edwinseng001@suss.edu.sg	<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.</p>

31	The Prophecy of Data Scientist Remunerations	Mr Victor Yiew victoryview002@suss.edu.sg	<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 Victor Yiew at victoryview002@suss.edu.sg with your CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>
32	Prognosis of Chronic Kidney Disease in Singapore	Mr Victor Yiew victoryview002@suss.edu.sg	<p>National Kidney Foundation (NKF) have more than 300,000 people suffer from chronic kidney disease and warned there could be even more undetected cases. The number of NKF patient inclusive haemodialysis and peritoneal dialysis (PD) on definitive dialysis increased to 23.7% from 7,007 to 8,668 from 2017 to 2021. An average of six people is diagnosed with kidney failure every day, which is threefold increase from two decades ago.</p> <p>This increase in numbers “causes major concern”, which NKF urge more to come forward for screening, importance of screening especially high-risk groups such as those with diabetes, renal failure cases and ageing elderly. Screening can detect early stage of kidney disease or organ failure. The disease worsens with kidney shrinking losing its ability to filter out waste product from patient’s blood. At final stage, the patient would require lifetime dialysis or a kidney transplant.</p> <p>On many occasions, people don't realise signs and symptoms of renal failure until it is too late, which is WHY they need to be educated, and public awareness be made to them, that early detection will help arrest or even severely delay the consequence of end stage renal disease.</p> <p>SUSS students can consider using IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software to uncover and prediction on early stage renal disease (ESRD) using the https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites for insightful and situational measure to advance on SG lifestyle.</p> <p>Those interested in this project please contact Victor Yiew at victoryview002@suss.edu.sg with your CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>

33	Diagnostic of diabetes ketoacidosis reversal treatment	Mr Victor Yiew victoryview002@suss.edu.sg	<p>Singaporean Diabetes can be reversed, even in people who have had it for several years was published by The Straits Times on 06 Aug 2022. Biochemically, Diabetic ketoacidosis (DKA) is defined as an increase in the serum concentration of ketones greater than 5 mEq/L, serum (blood) bicarbonate level is less than 18 mEq per L, blood sugar level greater than 250 mg/dL (although it is usually much higher), and a blood (usually arterial) pH less than 7.3.</p> <p>A study in Britain has proven that, and now, Singapore's National Healthcare Group (NHG) on simulate treatment can also work for Asians. The British study, called Direct, or Diabetes Remission Clinical Trial, had found that the majority who were able to achieve the target weight loss were no longer diabetic i.e. weight loss is the strongest link to remission, more than half of diabetics were overweight or obese. At the end of the first year of the study, 46 per cent were no longer diabetic. However, not all were able to maintain this, with only 36 per cent staying free of diabetes at the end of the second year. It is doing so through the Diabetes Reversal Programme that enable at least some diabetics to become free of a disease that can lead to severe consequences such as blindness, kidney failure and even death.</p> <p>SUSS students can consider using IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software to exploit DKA treatment 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 lifestyle habits, and government schemes and support to Diabetes-related ketoacidosis in Singapore.</p> <p>Those interested in this project please contact Victor Yiew at victoryview002@suss.edu.sg with CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>
----	--	--	--

34	Research Study on ChatGPT: Peek into Our Future	Mr Victor Yiew victoryview002@suss.edu.sg	<p>Its launch has led to widespread excitement. Its magic is that it democratized artificial intelligence (AI), enabling anyone anywhere to explore and play around with AI for the first time. There is much discourse about how it can free up valuable time from many job functions, so that people can focus on more priority and innovative tasks.</p> <p>ChatGPT has made it possible for people with limited technical expertise to interact with technology in a simple and accessible way. Looking back on the development of ChatGPT, creating an AI model that could understand and respond to natural language was a complex and technical process that was typically reserved for researchers and experts in the field. With its cutting-edge ability to generate human-like text, ChatGPT is the latest addition to the highly visible world of generative AI, which refers to AI systems that are designed to generate new and original content, rather than just recognizing or categorising existing data. These technologies are already advanced enough to transform every job and organization to help companies and individuals to removing mundane, repetitive and time-consuming tasks and becoming more productive.</p> <p>SUSS students can consider using IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework on innovative ChatGPT application in work and/or education to be used or produced and/or consolidating with data from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to associate and classify and/or develop analytical techniques to benefit future users.</p> <p>Those interested in this project please contact Victor Yiew at victoryview002@suss.edu.sg with CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>
35	The Prediction Analysis in luxury watch collection	Mr Victor Yiew victoryview002@suss.edu.sg	<p>An interesting up-rising demand for luxury watches and why collectors turn to the grey market in Singapore. Technavio published that luxury watches market size will increase by USD 1.78 billion with APAC holding 32% of the market share.</p> <p>Technavio forecasts the global luxury watch market will witness an incremental growth of USD 1.78 billion between 2020 and 2025 and expand at a CAGR of 1% during the anticipated period. The growth of these emerging markets is majorly driven by factors such premiumization through product design and innovation. In addition, other factors on these growing inclination toward fashion trends and the strong brand positioning and branding to influence the purchase decisions of customers will have an accelerating effect on the growth of this niche market.</p> <p>SUSS students can consider using IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework to predict on price incremental investment value from a legitimate source, https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to accolade the useful sources of information to develop useful algorithms for savvy investors to possess the finest collection of price-valued timeless watches.</p> <p>Those interested in this project please contact Victor Yiew at victoryview002@suss.edu.sg with CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>

36	The Prevention of Crypto Scam in Singapore	Mr Victor Yiew victoryview002@suss.edu.sg	<p>The global number of cryptocurrency scams has tripled yearly with both young and old victims lured by the prospect of simply investing to become rich quickly. This cryptocurrency scams had increased by 335 per cent in the first half of this year from the same period in 2022. In the first six months of 2022, it had been identified more than 2,000 domains register to promote cryptocurrency scams.</p> <p>In Singapore, this number of cryptocurrency scams reported to the police has jumped fivefold since 2019 with 631 reports made in 2021. These scams use videos containing deepfakes of popular personalities such as tech tycoon Elon Musk, Singapore actors luring victims into purported cryptocurrency giveaway in exchange for their crypto-wallet details and personal information. Does this advertising sound familiar on Facebook and other social media?</p> <p>SUSS students can consider using IBM SPSS Modeler, RStudio, Tableau Software, Power BI, Spreadsheet Modeling or other programming software and CRISP-DM framework to collate dataset from legitimate sources, https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to derive significant advantage and actionable value proposition.</p> <p>Those interested in this project please contact Victor Yiew at victoryview002@suss.edu.sg with CV by 01 Jun 2022 (cc JessTanwc@suss.edu.sg).</p>
37	Airbnb price & trend analytics	Mr Zhang Shuai sazhang001@suss.edu.sg	<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.</p> <p>For this project, 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. 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. Students who are strong/comfortable with Python are preferred.</p>
38	Fraud detection	Mr Zhang Shuai sazhang001@suss.edu.sg	<p>Twitter is a service for friends, family, and coworkers to communicate and stay connected through the exchange of quick, frequent messages. People post Tweets, which may contain photos, videos, links, and text. These messages are posted to your profile, sent to your followers, and are searchable on Twitter search.</p> <p>Similar to all platforms, there are lots of bot account instead of human account to tweets on the platform, which is supposed to be removed. In this project, dataset is provided with well labelling (human or robot).</p> <p>For this project, 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. Students who are strong/comfortable with Python are preferred.</p>

39	Classification of Economic Status based on Technology Development	Mr. Liew Sing Loon slliew001@suss.edu.sg	<p>A country's economic state depends a partly on its technology development. One of the measures of technology development is the number of patents filed.</p> <p>This project aims to build a classification or predictive model to predict the a coutry's economic status of countries by its technology development and related factors. The output of the model could be several classes, ranging from matured, emerging or declining economy. Such a model could potentially help policy makers to identify long term growth factors for their country.</p> <p>Students can consider using IBM SPSS Modeler, JMP, Tableau Software, Power BI or Spreadsheet Modelling to uncover the relationships.</p> <p>Some patent statistics can be obtained from: https://www.wipo.int/ipstats/en/index.html#data</p>
40	Learning Analytics Project	Ms Li Jizhi jzli002@suss.edu.sg	<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>
41	Customer Analytics	Ms Tang Yoke Wah ywtang001@suss.edu.sg	<p>Customer Analytics is the process of capturing and analyzing customer data to make better decisions; where analyses are based on customer behaviour and transaction. Building successful customer relationship through customer understanding and customer classification to identify potential profits and growth opportunities. Main CRM activities include Customer Acquisition, Customer Engagement and Customer Retention.</p> <p>Students are expected to deep dive into the phases of CRM, from how customers interact with the business, to generating insights through data mining, to gaining knowledge about customers, for the ultimate goal of designing strategies and planning for business growth and sustainability.</p> <p>Students are free to use any data mining tool /visualisation tool; or programming software for R or Python</p>

42	Making sense of IoT data	Ms Tang Yoke Wah ywtang001@suss.edu.sg	<p>IoT data is everywhere, captured by sensors all around the world we are in now; in elevators, on the road, in your smart watch, on the bus, etc.</p> <p>IoT data, in its raw state is basically meaningless. Much science and art goes into making sense of IoT data. An example is data captured by sensors installed in lifts. Every trip is recorded, and each day for each lift, there can be as many as a million data points. Such IoT data has volume, velocity, variety but highly questionable veracity.</p> <p>This project will require extensive data preparation: data engineering, data restructuring, and data cleaning that eventually leads to achieving the business goal. In the example of the lifts, the biz goal is to predict a lift's downtime (when will it next breakdown).</p> <p>Feasibility of project largely based on data availability that student can source for.</p>
43	Time Series Forecasting for Stock Indices or Forex Rates	Ms Yeo May Peng mpyeo002@suss.edu.sg	<p>Time Series Forecasting for Stock Indices or Forex Rates</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.</p> <p>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.</p> <p>Evaluate forecast results from the model and propose ways to make improvements.</p> <p>For students who are proficient in Python.</p>
44	Text Mining on Indonesian Text	Mr. Oh Chin Lock cloh001@suss.edu.sg	<p>In this project, we will perform text mining on Indonesian text to enhance and build predictive models of various activities such as e-commerce, banking/finance, healthcare, security services in the Indonesian market. Students will be expected to use a web scraping tool to collect Indonesian text for the relevant chosen sector, and then perform standard text mining tasks such as sentiment analysis, topic modelling and/or categorisation of the text. The topics, sentiments and categories will be used to build and enhance predictive models for the selected sector/problem. The insights will be used to enhance understanding of the various Indonesian sectors selected.</p> <p>Students will be expected to tune or obtain Bahasa lexicon databases from the public domain to enhance the text mining process. R, Python and/or SPSS modeller can be used for this project.</p> <p>More adventurous students can also try advanced neural net language models and techniques for this project.</p>
45	Using Analytics to understand Skincare products and preferences in Asia	Mr. Oh Chin Lock cloh001@suss.edu.sg	<p>In this project, we will collect (through web scraping) and analyze skincare products and services of popular skincare brands (e.g. Kiehl's, Shiseido, L'Oréal, K-beauty brands, etc), and compare them in meaningful visualizations. Existing skincare products and ingredients datasets can also be used to augment the data scrapped from the web. Students will also collect data about consumer reviews and preferences of these brands/products from relevant websites to understand key sentiments, and gaps where existing products fail to meet.</p> <p>These insights will potentially be used to inform the creation of a new skincare analysis product/service in the Asian market.</p>

46	Text Mining and/or Machine Learning on Sarcopenia Research (Healthcare)	Mr. Oh Chin Lock cloh001@suss.edu.sg	<p>Sarcopenia (loss of muscle mass) is a serious problem with the elderly. It is understood to begin when a person is in their 40s. Evidence suggests that skeletal muscle mass and strength declines in a linear fashion with up to 50% loss by the time a person is in their 80s. This is one of the most important causes of functional decline and loss of independence in older adults, with possible adverse outcomes such as disability, frailty, fatigue, falls and mortality.</p> <p>In this project, we will use text mining to understand the research related to sarcopenia done over the past 10 years. Students will be expected to retrieve sarcopenia related papers from research databases and to convert them to a format suitable for text mining. We will study the trend of the research. Multiple topic models and visualizations would be built to understand the progress in the field in the areas of clinical assessment, diagnosis, intervention and prevention researched all over the world. At the end of the project, students will present insights related to sarcopenia research and the most recent developments, particularly in prevention. Existing structured datasets (one e.g. https://doi.org/10.1016/j.dib.2022.108881) can also be curated/merged to build a predictive tool for easy assessment of sarcopenia.</p>
47	NLP with Generative AI on Supply Chain News	Mr. Oh Chin Lock cloh001@suss.edu.sg	<p>This is a possible collaboration with ASEAN Applied Research Centre (AARC), a non-profit organization hosted at Supply Chain City in Singapore. Students will be expected to collect or scrape data from reputable supply chain related news sites (e.g. Supply Chain Dive, Supply Chain Brain, CSCMP, etc) and perform NLP/Text Mining on these news to understand the key topics, trends and developments in supply chain affecting various sectors. In addition, students can also explore the use of Generative AI to assist data collection, analysis and summarization, and to compare the insights from using different techniques. A final report may be presented to stake holders at the AARC at the conclusion of the project.</p>