ANL488 Project List for Jul 2021 Semester

No	Project Title	Description
1	Resource Optimisation for processing of resale applications (Supervisor: A/P Ma Nang Laik)	The student will work with the Housing & Development Board (HDB) to develop analytical models for resource optimization to meet the business objective of completing the resale applications within 8 weeks from the date of registration. The student needs to understand the business process of resale flat application and identify the bottleneck by interviewing the staff involved in the process during this project. HDB will provide anonymised data for the study. The student will use the relevant analytics tools to assist HDB to meet the business objective based on a finite pool of resources and recommend the resource requirement if there is a shortfall in manpower. An accurate model will help HDB to optimize its manpower and counter resources. The student needs to be good at mathematics and have some basics programming skill to do the optimisation engine. Those interested should submit their CVs to A/P Ma (nlma@suss.edu.sg) and A/P Lee (leeyh@suss.edu.sg) by noon on 4 Jun 2021. Selection will be very competitive.
2	Modelling and Forecasting Time Series (Supervisor: Dr Karl Wu)	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. 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. Those interested please email Dr. Wu at karlwuky@suss.edu.sg by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).
3	Analysing zero-inflated data with R	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

	(Supervisor: Dr Karl Wu)	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. 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 7 Jun 2021 and cc (leeyh@suss.edu.sg).
4	Pricing Analytics for Food and Beverage Business (Supervisor: Dr Liu Wenting)	This project requires the student to work with Revenue Management Solutions (a consulting firm focused on data-driven pricing strategies and data analytics for Food & Beverage business). This project will focus on the Food & Beverage business. The major purpose is to analyse the existing data hosted by the company, e.g., POS data, consumer satisfaction data or demographics, and optimize pricing, make recommendation to menu design, provide analytical support to marketing and financial reporting which collectively deliver a profit gain. Hence, the project will involve qualitative and quantitative analyses, with the objective of delivering set of recommendations in one of the key strategic area – value-based pricing, operation analysis, customer segmentation, or competitor insights. 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 7 Jun 2021. Selection will be competitive. Those working in the same industry as this company will not be selected for this project to avoid any conflict of interest.
5	Advanced analytics to assess Corporate Risk and predict default (Supervisor: Mr Chua Poh Chai)	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. 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. 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.

		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. Students who are keen in this project can email Chua Poh Chai at pcchua002@suss.edu.sg by noon 7 Jun 2021 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.
6	Learning Analytics: Predictive Model on at-risk students (Supervisor: Mr Edwin Seng)	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. 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 7 Jun 2021 and cc (leeyh@suss.edu.sg).
7	Econometric model building with regression analysis (Supervisor: Dr Tan Khay Boon)	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. Students should be familiar with statistical software such as JMP, SAS or EViews which can perform regression analysis. Those interested in this project please contact Dr Tan Khay Boon at kbtan010@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).

8	Forecasting using time series data (Supervisor: Dr Tan Khay Boon)	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. Students should be familiar with statistical software such as SAS or EViews which can perform regression and forecasts. Those interested in this project please contact Dr Tan Khay Boon at kbtan010@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).
9	Modelling, analysis and regression based on time-series data (Supervisor: Dr Chris Ho)	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. Please email Dr. Chris Ho (jhho003@suss.edu.sg) directly to express your interest in working on this project by noon 7 Jun 2021and cc (leeyh@suss.edu.sg).
10	Product recommendations based on sentiment analysis/natural language processing (Supervisor: Dr. Chris Ho)	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. Please email Dr. Chris Ho (jhho00@suss.edu.sg) directly to express your interest in working on this project by noon 7 Jun 2021and cc (leeyh@suss.edu.sg).
11	Customer insights analytics (Supervisor: Zhang Shuai)	Customer behavior analysis is not conducted regularly nowadays. A huge amount of time is needed for manual data extraction & preparation to understand customer behaviors. The lack of information may cause the following issues such as 1. Lack of customer buying pattern visibility, 2. Lack of customer growth, churn & revenue impact trends/knowledge, 3. Difficulty in developing the right strategies to tackle business challenges. This project aims to analyze customer data to identify customer behaviors to drive growth. With the proper study of customer behaviors, the company can prioritize their business strategy to facilitate business growth. The analytics result can be derived (no limit) from the following aspects, customer churn, credit analytics, customer segmentation, etc. This project will evaluate students' capability from different areas, including business case studies, exploratory data analysis, machine learning techniques (Random forest, decision tree, regression, classification, etc.) In the end,

		students are expected to provide suggestions to the management team on the business direction based on analytics results. 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 7 Jun 2021and cc (leeyh@suss.edu.sg).
12	Turning Social Media Conversations into Insights (Supervisor: Mr Adam Wong)	Twitter is a microblogging and social networking service where users post content and interact with posts known as "tweets". Today, Twitter has popularised the use of hashtags as a way to group conversations and allow users to follow conversations on particular topics. Every second, on average, 6,000 tweets are tweeted on Twitter. With over 500 million tweets per day and around 200 billion tweets per year, you can imagine how rich with information this platform is. The objective of this project is to gather and analyse publicly available twitter data in order to discover interesting information and hidden patterns (or to deliver the earliest signals for breaking news, real-world events, off the radar context and perspective, and emerging trends – for example, #bitcoin). Students should be familiar and comfortable in using R programming for the collection, cleaning, preprocessing, modelling and analysis of textual data. Those interested in this project please contact Mr Adam Wong(adamwong002@suss.ed.sg) by noon 7 Jun 2021 and cc(leeyh@suss.edu.sg).
13	Text Mining on Qualitative Survey Responses (Supervisor: A/P Lee Yew Haur)	Two students will work with SUSS to use text mining on the qualitative responses to the Graduate Employment Survey 2020. Each student will tackle 2 qualitative questions each. 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. Those interested should submit their CVs to A/P Lee Yew Haur(leeyh@suss.edu.sg) by noon on 7 Jun 2021. Selection will be competitive.
14	Analysis of Healthcare related Patents filed during Covid Pandemic in 2020 (Supervisor: Mr Oh Chin Lock)	Since the advent of the Covid pandemic in late 2019, what are the healthcare-related patents filed in WIPO (World Intellectual Property Organization) under PCT (Patent Cooperation Treaty) filing? This project will look at the Google Patents public dataset to analyze and visualize healthcare patents filed in 2020 that are related to COVID. The analysis will provide a useful patent landscape to understand the intentions of companies and even perhaps insights into the strategies of countries in covid patent control over the next few years.

		As part of the project, you'll learn about international patent filing and how companies protect and guard their markets through patents. Students can use IBM SPSS modeler or Python text analytics libraries to analyze/visualize the patent database, and use a public patent search tool such as Google Patents to extract relevant data. Useful references: * WIPO Patent Statistics: https://www.wipo.int/edocs/infogdocs/en/ipfactsandfigures2019/ * What is PCT filing? https://www.wipo.int/pct/en/faqs/faqs.html Those interested should submit their CVs to Mr Oh(cloh001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon on noon 7 Jun 2021.
15	Predicting Student Performance (Supervisors: Mr Adam Wong, Ls Li Jizhi & Mr Prasanna Rao)	Supporting student learning has always been a critical part of higher education, especially so for adult students who face more challenges and multiple commitments for their time. Hence, support from Institutes of Higher Learning (IHLs) can play a pivotal role in helping the adult students towards achieving greater academic success. Support for student can be provided after their performance has fallen below a threshold. Or, IHLs can leverage predictive analytics to identify students who have a higher possibility of under-performance and help them before it happens. The objective of this project is to identify possible determinants of academic performance and use this information to construct predictive models to classify (or predict) under-performing students apart from good-performing students – from an existent dataset. Learning support and intervention can then be provided in a more timely and targeted manner to help students succeed. Students should be familiar and comfortable in using IBM SPSS Modeler. Students may propose alternative tools/software after discussing with your supervisors. This project will accept 3 students. Each student will have a different supervisor. Those interested in this project please contact the respective supervisors by 7 Jun 2021 and cc(leeyh@suss.edu.sg): Mr Adam Wong at adamwong002@suss.ed.sg Ms Li Jizhi at jzli002@suss.edu.sg Mr Prasanna Rao at prasanna001@suss.edu.sg

16	AI predictive modeling of the business expansion of a Singapore logistics company into Southeast Asian countries (Supervisor: Dr How Meng-Leong)	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: (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). 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. 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. 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 (shawnhow001@suss.edu.sg) and cc(leeyh@suss.edu.sg) by noon 7 Jun 2021.
17	Detecting Accounting Frauds in public listed firms (Supervisor: Dr Shi Lirong)	The consequences of accounting fraud are severe. However, accounting frauds are difficult to detect. This project aims to develop effective models which can predict accounting fraud on a timely basis out of sample. The project will investigate various machine learning techniques to build fraud prediction model, and use various performance evaluation metrics to select the best prediction model. Students who are keen in this project can email Dr Shi Lirong at lshi@suss.edu.sg by by noon 7 Jun 2021 with their CV and cc(leeyh@suss.edu.sg). Accounting or finance knowledge is not mandatorily required but statistics and machine learning, together with strong Python/R programming will be helpful.

18	The reliability of Renewable Energy to replace oil and gas as our energy of choice (Supervisor: Dr Munish Kumar)	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. 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. 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 7 Jun 2021 and cc (leeyh@suss.edu.sg).
19	The Evolution of Popular Tunes (Supervisor: Dr Munish Kumar)	For over a century, it has been documented that music reflects society and community. In studying the evolution and way of life of tribes and people, anthropologists explore the music they make and the words they use in their songs. Thus, to postulate that people like to express their state of being through songs is no stretch of the imagination. Using NLP and techniques associated with data mining, analyse the top 20 most popular songs over each 10 year era from 1970 to 2020 (you will have 100 songs). Create a visual map of the words used in the songs; you will need to then need to correlate the words with the events in those eras and determine, if indeed, songs were reflecting the state of mind of the people. 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 7 Jun 2021 and cc (leeyh@suss.edu.sg).

20	The Prevalence of Stroke Analysis Using Data Science (Supervisor: Mr Victor Yiew)	Singapore Statistics has issued that every day 17 people die from cardiovascular disease (heart diseases and stroke) accounted for 29.3% of all deaths in 2019, which is almost 1 out of 3 deaths due to heart diseases or stroke in Singapore, this is published on 20 August 2020. This study on stroke as a neurological deficit attributed to an acute focal injury of the central nervous system (CNS) by a vascular cause, is the leading cause of adult disability and fourth most common cause of death and accounted for more than 10 percent of all deaths - burden of stroke will rise with ageing population. The student will use the IBM SPSS Modeler to predict whether a patient is likely to get stroke or conduct survival analysis to the time taken for an individual to reach the certain event using the viable parameters from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to accolade all useful emergent discoveries to an early societal precautionary awareness programme. Those interested in this project please contact Victor Yiew at victoryiew002@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).
21	Market Research on Organic Products (Supervisor: Mr Victor Yiew)	An understanding of a healthier world awaits us relates to belief of consuming nutritious food, mind and body workout and breaking environmental pollutants. Organic food products are grown and processed using organic farming practices, which encourages soil and water conservation as well as reduction of pollution. Supermarket chain FairPrice carries more than 800 organic items, ranging from fresh produce to household items, up from 200 since 1998. As local consumers become more affluent, they are better able to choose what they eat. Using the customer loyalty program to incentivize initial buyer coupon on organic products to all the loyalty program participants towards collective purchases that could classify whether selected customers has purchased any of the organic products. The student will use the IBM SPSS Modeler to determine which customers are most likely to purchase the organic products through survey or consolidating of data from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to classify and/or develop prediction for organics purchase Indicator to offering new and existing customers a new line of organic products or future campaigns. Those interested in this project please contact Victor Yiew at victoryiew002@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).

22	Data Visualization on Fertility Rate through Story Showing (Supervisor: Mr Victor Yiew)	In 2020, the fertility rate for Singapore was 1.2 births per woman. This fertility rate has fell gradually from 3 births per woman in 1971 to 1.2 births per woman in 2020. Singapore offers 'pandemic baby bonus' to boost births: it's a one-off payment to encourage people to have babies during the coronavirus pandemic since Singapore's fertility rate is looking up on positive marriage trends this year. Total fertility rate represents the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with current age-specific fertility rates. Singapore has one of the lowest birth rates in the world, which has been a struggled to boost for decades. This is time to embrace on your ability to build something significant, meaningful to life, to answer a question you care so much for, at the same time to discover more on holistic government support and incentives around you. The student will use the IBM SPSS Modeler, RStudio, Tableau Software or other programming software to exploit the boost of Births and Fertility in Singapore from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to develop on Effectual Infographics through Story Showing on socio-demographic, environmental factors, health status, mental preparation and life habits, and government schemes and support for bearing a healthy baby in Singapore. Those interested in this project please contact Victor Yiew at victoryiew002@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).
23	The Fun Side of Data Science (Supervisor: Mr Victor Yiew)	Singapore is strategically location at the heart of South-East Asia, well-developed with a unified and vibrant cultures, safety with efficient infrastructure and wide variety of attraction makes it an attractive tourism destination. The international tourism has accounted for 4.1% of Singapore's national GDP in 2017, unswerving contribution of $17.7 billion. The percentage of tourism's direct contribution to Singapore's GDP is projected to rise to 4.4% in 2028, the number of international arrivals at 18.5 million dwarfs the resident population of 5.64 million. Is our SG Digital Voucher 2020 adequately replicate the indirect contributions to this projected GDP? Singapore Sports Hub National Open Championships 2021 continue with momentum? With an advanced vaccine development and speedy distribution be suffice to the revival of the global economy? It will be interesting to apply data analytics to visualize those significant factors and beneficial to our tourism sector to shape a sustainable economy. The student will use the IBM SPSS Modeler, RStudio, Tableau Software or other programming software to uncover and have fun and focus on professional goals for this project using the https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to translate association and cluster on heritage tours, renowned attractions, hotel staycation, cruise destination and more to predictive analysis on popular attractions favorable to boost beyond the Singapore GDP.

		Those interested in this project please contact Victor Yiew at victoryiew002@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).
	Emerging Online Datafication on the Consumer Behaviors (Supervisor: Mr Victor Yiew)	Singapore and millions of people all over the world are forced to change our lifestyle habits during the COVID-19 pandemic. Challenges arising from this outbreak had accelerated the use of existing and new technologies and tools that a new digital spending norm has emerged. The six key consumer behavior threshold levels has identified an early signals of spending patterns, particularly for emergency pantry items and health supplies, and these patterns are being mirrored across multiple markets. Consumer ease of convenience and accessibility have increased existing online spending as well as converted many offline shoppers to online. This trend is unlikely to return to the same levels of online shopping before the outbreak. This insights to these new digital spending trends could be useful to eCommerce strategy, resources and logistics design blueprint on the 5G technology and network integration. https://www.bestinsingapore.co/best-online-shopping-sites-singapore/ and https://data.gov.sg/dataset/online-shoppers?resource_id=3038ccbd-a78e-4fbe-a9a4-fde9230f480a The student to use the IBM SPSS Modeler and CRISP-DM framework to examine qualitative dataset from https://data.gov.sg or https://www.singstat.gov.sg/ or approved sites to derive on insightful and actionable value proposition. Those interested in this project please contact Victor Yiew at victoryiew002@suss.edu.sg with your CV by noon 7 Jun 2021 and cc (leeyh@suss.edu.sg).

Updated on 1 June 2021