**Shang-Chin (Jonathan), Lee**

• 979-267-1233 • [scleeza@tamu.edu](mailto:scleeza@tamu.edu) • [LinkedIn](http://www.linkedin.com/in/scleeza)• [Website](https://scleeza.github.io/portfolio/)

**Python | SQL | Data Analytics | Machine learning**

|  |
| --- |
| **EDUCATION** |
| **Texas A&M University (TAMU)**, *College Station, Texas. May, 2020*  Master of Engineering in Industrial & System Engineering.  **Relevant Courses:** Survey of Optimization, Machine Learning and Data Analysis, Computational Tools and Database in Big Data, Simulation Modeling and Applications, Design of Experiment. |
| **National Taiwan University (NTU)** ,*Taipei, Taiwan. Aug, 2013*  Master of Science in Applied Mechanics. |
| **Chang Gung University**, *Taoyuan, Taiwan. Aug, 2011*  Bachelor of Science in Mechanical Engineering. |
| **SKILLS** |
| **Programming Languages:** SQL, Python, C++/C#, Bash/Linux, MATLAB.  **Data Science Library**: Pandas ,NumPy , Scikit-learn, TensorFlow/ Keras, Spacy, Matplotlib, Spark, SciPy, BeautifulSoup.  **Tools:** Microsoft Excel, Tableau, Git.  **Domain Knowledge:** Hypothesis Test, Regression models, Supervised/Unsupervised learning, NLP/NLU, DOE (ANOVA, A/B test), Time Series Analysis, Linear and Non-Linear Programming, Inventory control, Cost Analysis, Quality Control. |
| **EXPERIENCE** |
| **Taiwan Semiconductor Manufacturing Co., Ltd.** (TSMC), *Taiwan. Oct 2013 – Jan 2017*  *Equipment Engineer* (Full-time)   * Identified abnormal patterns in throughput by extracting WIP data from ERP database and connected with spreadsheet written with analysis pipeline by VBA to detect bottle neck and conducted root cause analysis over 50 process. * Collaborated with cross-functional team to complete throughput improvement project by saving USD 0.5M and increased 6% throughput, providing data analytics reports, experiment results, resources estimation. * Forecasted future demand of spared parts through daily and monthly consuming rate and built ARIMA models on predicting demands to maintain daily operations and increased cost-effectiveness. * Managed 3 KPIs (Throughput, Parts Consuming, Cost) for 2+ years, generating weekly analysis reports to stakeholders with data driven insights and engineering solutions to keep continuous improvement.   **Chang Gung University**, *Taiwan. Aug 2017 – Jun 2018*  *Research Assistant* (Full-time)   * Lead a task team to develop a prototype of portable ultrasound scanner, collaborated with hardware suppliers and physicians to advanced system’s user experience and deployed prototypes into 10 hospitals. * Implemented images analysis algorithm using convolution kernels to classify severe level of fatty liver written by C++ (OpenCV/OpenMP), and successfully raised funds from investors for over $USD 60K.   Created a GUI for our system and store 1000+ trial testing results in a MySQL database for further query.  **Social Impact Analytics Institute (NGO),** *Seattle, WS. Aug 2020 – Present*  *Data Scientist*   * Reduced 70% work load of group’s researchers by generating pdf files process and visualization streamline on python scripts using technique of word cloud and LDA, model was trained from corpus, using Spacy and Gensim libraries to do text cleaning, tokenization and clustered into correlated topics with its word cloud. * Applied data mining on text files to extract information like time, location, names by NER tag using SpaCy and have them as new keywords to increase web scrappers efficiency. * Deployed 10+ web scrappers to collect demographic data or documents from public databases or websites. |
| **PROJECTS** |
| **LSTM in Time Series Analysis**   * Deployed RNN/LSTM models on a streamlit dashboard to predict covid-19 case in worldwide, models were trained using TensorFlow, outperformed traditional ARIMA model 90% of MSE in validation stage.   **Customer Questionnaire Analysis**   * Analyzed datasets from 1000+ customers questionnaires, using Scikit-learn library to do KNN imputation to fill MAR missing values with observations have similar behavior patterns. * Trained models to predict whether people would stay after free trials, using Scikit-learn to do grid search with multiple models and found feature importance by random forest and linear regression model.   **Team Mate Recommendation**   * Created a team mate suggestion function for TAMU datathon website, using Scikit-learn library to preprocess 500+ questionnaires datasets and then clustering similar participants by K-mode algorithm |