|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Huy Nguyen (James)**viethuy2108@gmail.com5147469798[**https://www.linkedin.com/in/huy-nguyen-04782414a/**](https://www.linkedin.com/in/huy-nguyen-04782414a/)  |  | | --- | | **Github**[**https://github.com/viethuy25**](https://github.com/viethuy25) | | |  | | --- | | Summary A Data Analyst and IT Security Enthusiast with a Software Engineering background | | **Skills and qualifications**  * **Operating Systems:** Windows, Linux * **Programming language:** HTML, JavaScript, CSS, Python, SQL * **Framework and Data Modeling:** Spark, Pandas, NumPy,   Sckit-Learn   * **Data Visualization:** Seaborn, Matplotlib * **Machine Learning:** Linear Regression, Bayes Classification, K-Clustering Classification, Six Sigma * **Security knowledge:** TCP/IP, firewalls, routers, and network protocols and technologies * **Tools:** Jupyter Notebook, Jupyter Lab, Kubernetes (studying) * **MS Office skills**: Word, Excel and PowerPoint * **Language:** Vietnamese and English – Fluent * **Certification:** CompTIA Security+ Certified- June 2020 * **Soft skills:** Communication, Teamwork, Curiosity, Adaptability, Self-driven, Attention to Detail, Problem Solving, Commitment | |
| **Education**Bachelor of Computer Science – Information Systems June 2020Concordia University, Montreal, Quebec | **Volunteer Experience or Leadership** **Tutor Sep 2015- May 2016**  Bronte College, Mississauga, Ontario  Role: Tutor math and science for grade 9 to 11 students  **Participant at Terry Fox Run Sep 2015** |
| **EXPERIENCE AND PROJECTS**  * [**NBA stat analysis**](https://github.com/viethuy25/nba_stat) **(Personal) July 2020- On going** * Display data overview and perform **data cleansing**. * **Graph height and weight distribution** and display the change over time with **scatter chart**. * Analyze **draft round and draft numbers correlation to players performance** over seasons **with heatmap and scatter chart**. * **Data mining** from stats.nba.com and basketball-reference.com. **Clean and reformat** csv and dataframe. * **Hosting a** **web application** and **provide data-driven** **prediction** impact of players on **AWS Cloud** (**planning**). * *Language used: Python, NumPy, Panda, Matplotlib, Seaborn, Plotly, nba-api, S3 Buckets, AWS.* * [**Bank loan category analysis**](https://github.com/viethuy25/bank-loan-prediction) **(Personal) Jan 2020- Sep 2020** * Display data overview and perform **data cleansing**. * **Graph bar chart** to show overview number of loans per category. * Perform **Naïve Bayes and K-mean clustering** models with and without **emp\_title** to compare **accuracy scores** to verify **weight of emp\_title** in approving loan is **minimal**. * *Language used: Python, NumPy, Panda, Matplotlib, Seaborn, Plotly.* * **Bookstore website (Academic) Sep 2019- Dec 2019** * A bookstore website that is able to **search, view, add, delete books in database**. * My role: back-end development, unit testing, redirect unfound extension DNS, design book object, auditing, sql-injection test, cohesion & coupling design, exception handling . * *Language and library used: JavaScript, HTM, SQLL.* * [**Yelp! Big Data Analysis**](https://github.com/nguyenthanhtung2605/yelp-data-analysis) **(Academic) March 2019- April 2019** * Used **Kaggle public database**, applied **user-user collaborate filtering and frequent itemset** algorithm with Pyspark and give data-driven recommendation based on user’s inputted postal code * Use Pyspark Dataframe method to **calculate RMSE and MSE of result recommendations**, suggest possible improvement and report result * *Language and library used: Python, PySpark.* | |