Samuel Tseng

0177 3203241   samueltseng@icloud.com   Alice-Berend-Straße 3, 10557 Berlin

Looking for a full-time Data Scientist / Machine Learning Engineer position in Berlin. To leverage my data related experiences and skills into your company.

• 1st Prize from Berlin venue and Top 10 Winners of the Global Voice & Travel Hackathon 2019, works with Mia Chang and Muhammad Samir.

• Fluent in Machine Learning and Natural Language Processing related libraries in Python.

• Teamwork developer, lifelong learner and contributor, co-organiser of Taiwanese Professionals in Germany (TPG).

Experience

DATA SCIENTIST, DIGITTY.IO GMBH;
BERLIN, GERMANY – FEB 2019 - PRESENT
Tech stack: Machine Learning, Data Analysis, Web Scraping, RESTful API, Python, spaCy, Flask, Scrapy, MongoDB, Redis Google BigQuery, Google Firebase, Git, Docker, Jira.

1. dignity.io Brain API: Built RESTful API by using Python and Flask. API's Users are backend engineer inside digitty.io, the main function is to extract appropriate data from Google BigQuery and Google Cloud Firestore to calculate similarity between data and user's input. Embed machine learning model inside the API, for automate data classifying use.

2. Applicants' CV parser: Parse CV's text data then use spaCy to implement Name Entity Recognition (NER). Customise entity of applicant's CV for grabbing more useful data.

3. Job description web scraper: Built several web scraper for crawling job description of job posting, using Scrapy for main scraping process and MongoDB for backend database.

DATA CONSULTANT, FREELANCER;
BERLIN, GERMANY & TAIPEI, TAIWAN – DEC 2017 - PRESENT
Tech stack: Machine Learning, Python, Pandas, Scikit-Learn, Matplotlib, NLTK, Keras, Git, Docker, MongoDB.

1. Mandarin (zh-tw) News Corpus Dataset: Building a dataset by crawling news from several Taiwanese famous online news website, use Scrapy, MongoDB, Docker and Pandas to set up spiders and the pipeline for database. This project has still updated continually to increase new spiders, clean and extract text data needed.

2. Costumer's Convertion Rate Analysis: To realize what makes users convert, like buying a voucher when they get an offer. I create feature that extract from user and offer, then some features might have higher frequencies of vouchers' payment, and I wish to add that information so that calculate the confidence rate next. Using Extreme Gradient Boosting to create model, then print the feature importance by weight, gain, and cover to analysis which feature important is.

3. Sentiment Analysis: Training by the text from Twitter Corpus and news' data from BBC, CNN, etc. to analysis customer's convert ratio, based on Twitter Sentiment Analysis Training Corpus and clients' data, using Pandas, NLTK, scikit-learn, Scrapy modules and MongoDB database to achieve it.

Volunteer Experience

CO-ORGANIZER, TPG (TAIWANESE PROFESSIONALS IN GERMANY);
BERLIN, GERMANY — MAY 2019 - PRESENT

From LinkedIn network to offline meetup, we attempt to connect not only Taiwanese but every professionals in their specific area, to share and learn together. We have monthly Stammtisch and seasonally keynote speech or workshop. For now, we have 30 active members in Berlin.

TUTOR, CODEDOOR.ORG;
BERLIN, GERMANY — OCT 2018 - PRESENT

Be a Machine Learning mentor in this project held by CodeDoor, provide basic Machine Learning consultant. CodeDoor is a non-profit organization that provides refugees and migrants the opportunity to learn to code. Participants receive a scholarship backed by silicon valley companies and enter a network of like-minded developers and companies.

Education

POSTGRADUATE, TAIWAN AI ACADEMY, ACADEMIA SINICA;
TAIPEI, TAIWAN — JAN 2018 - APR 2018

The bootcamp held by Taiwan AI Academy, Academia Sinica. It covers Machine Learning and Deep Learning in depth through theory to practice. The course included from basic Python instruction to advanced, like NLP, image classification, etc. All learners in bootcamp can choose one or two topics to dig into if you want. By the end of the term, all participants need to build a team and access hands-on projects to deal with real issue from partner companies/organization.

- Group Project: Detect Human-Object Interaction - Chunghwa Telecom Co., Ltd

BACHELOR OF SCIENCE, APPLIED MATHEMATICS, UNIVERSITY OF KAOHSIUNG;
KAOHSIUNG, TAIWAN — SEP 2006 - JUN 2010

1. Probability & Statistics

2. Combinatorial Mathematics

3. Programming Language(C, C++, R)