linkedin.com/in/lei-c

LinkedIn

Lei Cao

lcao4@toromail.csudh.edu cell (310) 593-1930
Carson, CA 90746

github.com/leicao666

Github

LinkedIn: https://www.linkedin.com/in/lei-c/

Github: https://github.com/leicao666

Highlights: (This CV has Two pages)

- (GPA 4.00) MS Computer Science, CSU, Dominguez Hills
- 5 years as a Software Engineer, Data Science Intern, AI Intern
- Data Analyst Nanodegree from Udacity.com
- Mastering Data Science Applied Lab, TheDevMasters
- Experience in Internet Startup projects

Experience

Aritificial Intelligence Intern at Data Application Lab ……Feb 2018 – Present

- Create NLP class content with AI Bootcamp teacher
- Create a Chatbot using TensorFlow with word2vec
- Assist to setup a MongoDB server

Data Science Applied Lab at theDevMasters …… Oct 2017 – Present

- Project Based Learning
  - Recommendation Systems
  - ***Natural Language Processing and Sentiment Analysis***
    - Wrote algorithm, utilized Twitter API, search an topic
    - From scratch, analyzed tweets content into positive and negative sentiment
  - Big Data with Spark and Splunk
  - Deep Learning and Time Series
  - ***Computer Vision with OpenCV***
    - Used webcam to create a real-time facial&object recognition program
  - Created a web scraper on a Django platform.
  - Blockchain Technology, Aritificial Neural Nets

Data Science Intern at Yanset Career Success Platform (http://yanset.com/) ……Oct 2017 - Present

- Selenium + PyUnit End-to-end testing.
  - Using Python Unittest frameworks and Selenium API
  - Coding test cases to simulate user's actions on the company's website
- Web Scraping(Beautiful Soup) and Data Analysis

Graduate Research at California State University-Dominguez Hills ……Aug 2017 - Present

- ***Project: Big Data Analytics and Visualization of California Transportation***
  - Presented in HENAAC '17 Conference
  - Built a model to predict road accident trend given economic data
  - Built Apache Spark and Hadoop cluster from scratch in Chameleon Cloud
  - Learned: I. How to find valuable insights into large datasets, in this case, a correlation
    II. How to use Spark to perform big data analysis

Software Engineer at Meron English Edu. Science & Technology Co., Ltd. ……Jun 2012 - Nov 2016

- Project1: School Database Software System Development - Financial and Administrative System
  - VBA, Microsoft Access Database, SQL (3000 + lines of code, 100+ features)
  - Provide data science insight to assist advertising and recurring billing cycles
  - The software system handles 500 enrolled students per semester, 5000 overall.
- Project2: Virtual Reality iOS App ……Nov 2015 - Mar 2016
  - Immersive English Learning Experience using Panoramic Video
  - Software: Unity3D + Google Cardboard SDK + Cardboard goggles. Hardware: Ricoh Panoramic Camera

Skill Sets

- Data Science – Pandas, Numpy, Matplotlib, Tableau, Seaborn, Machine Learning, Deep Learning, Selenium, PyUnit
- Machine Learning – Natural Language Processing, Recommendation System, Image Classification
- Big Data Analytics – Spark Deployment, Hadoop cluster Deployment, Distributed Computing, MapReduce Algorithm
- Artificial Intelligence – Deep Learning, Convolutional Neural Network, TensorFlow, Softmax Regression, Machine Learning
- Programming Languages – Python, Java, SQL, VBA, C++, Linux, CSS, HTML, JavaScript, Scheme, Prolog, R
- Misc. – AWS Cloud Deployment, Chameleon Cloud, Jupyter Notebook, Microsoft Access Database, MongoDB
- Virtual Reality – C#, Unity3D, Google Cardboard VR video player development

(Projects on the Next Page)

Projects

Loan Status Prediction……..……..Jan 2018

- Used given bank customer info on loans to predict if a new customer is likely to fully pay the loan.
- Used multiple machine learning algorithms

Titanic Passenger Survival Prediction ……..Jan 2018

- Kaggle Competition Leaderboard Top 5%
- Machine Learning: Gradient Boost Classifier, 887 features,

House Price Prediction in King County ……..Jan 2018

- Machine Learning – Linear Regression,
- Feature Engineering

FBI Gun Purchase Background Check Data Analysis ……..Dec 2017

- Found a seasonal pattern for gun shopping
- Found the most related variable of US census data to FBI Gun data

Image Classification - MNIST Handwritten Digits……..Nov 2017 - Present

- (Accuracy 99.2%) Built A Convolutional Neural Networks in TensorFlow.

Hadoop Cluster in Chameleon Cloud ……Mar 2017 - Jun 2017

- MapReduce algorithm to compute and analyze a multi-year H1B visa petitions dataset.
- Visualization in Jupyter Notebook by Python Pandas and Matplotlib
- Learned: I. How to set up a Hadoop cluster in a Cloud environment from scratch.
  II. MapReduce algorithm, Visualization

Education

MS Computer Science, California State University-Dominguez Hills ……..**GPA 4.0/4.0** Expect Grad. Dec 2018

- Courses: Artificial Intelligence, Cloud Computing, Software Engineering, Data Structure, Algorithm, Object-oriented Design, etc.

Udacity Data Analyst Nanodegree ……..Nov 2016 - Present

- Courses: Descriptive/Inferential Statistics, Explorative Data Analysis, Data Wrangling, Visualization, Machine Learning

BA English, Anhui University - China ……..Sep 2003 - Jun 2007

- President, Spoken English Association

Volunteer

Data Library Website Evaluator for LA County – LA Counts (http://www.lacounts.org/) ……..Aug 2017

- Website Beta Evaluation and Testing
- Commented on Data Source quality, Search Scenarios, Data Visualization, Overall Experience, etc.

Other Activities and Awards

- Research Poster Competition in HENAAC 17 conference ……..… Oct 2017
- A member of ACM and CAHSI (Computing Alliance of Hispanic-Serving Institutions) ……..Since Sep 2017
- Municipal Coding Champion – Elementary School ……..1996
- Work Experience in Sudan for two years as an Interpreter & Business Manager …….. 2008-2009
- President of the Spoken English Association, Awarded "Excellent Student Organization Leader" ……..2004