ZHAOHUI (MARVIN) MAN

zhaohui@live.unc.edu | (984)2157418 | LinkedIN: https://www.linkedin.com/in/zhaohui-marvin-man-2832142b/

Github: https://github.com/MANZHAOHUI

124 St. Andrews Ln., Chapel Hill, N.C. 27517

EDUCATION AND CREDENTIALS

University of North Carolina at Chapel Hill

MSIS (Master of Science in Information Science)

CHIP (Carolina Health Informatics Program)

Graduate Certificate in Clinical Information Science

PLA University of Foreign Language, China

MA in English Language and Literature

BA in ETS (English for Science and Technologies)

SKILLS AND CERTIFICATS

Computer Skills: strong knowledge of Linux/UNIX-based systems, Python (Numpy,Pandas, Matplotlib, Seaborn, Scipy, Scikit-learn, Tensorflow, Keras, Nltk, Gensim, BeautifulSoup, etc.), R, SAS, Tableau, SQL, VISIO, OpenRefine, LucidChart, LightSIDE, Apache Spark, MySQL, Machine Learning toolkits, Access, Qualtric, Microsoft SQL Server, Bioinformatics toolkits (GSEA, GOrilla, shortBRED, Cell Ranger, BEDTOOLS, Metaphlan, Kraken2, Uniprot, Cytoscape, Scanpy, etc.)

Web Skills: HTML, CSS

Data Analysis Skills: Manage EHR(Electronic Health Record) , Build dashboard, Define business metrics, Machine Learning and Deep learning algorithms such as Linear Regression, Logistic Regression, Random Forest, SVM, KNN, CNN, RNN, and TensorFlow; NLP (Natural Language Processing); Sentiment Analysis

Certificates: TESOL (Teaching English to Speakers of Other Languages) Advanced Certification, TESOL Trainer Certification

Language Skills: Mandarin (native), English (proficient)

PROJECTS AND PARTS PLAYED

Human Lung Peptides Hits Detection, Mar. 2020- Jun. 2020

- Detect peptide hits using machine learning package DBCAN based on sequencing results
- Visualize the results through 2D and 3D scatter plots using python packages MATPLOTLIB and SEABORN
- Annotate data points with their unique identifiers in the 3D plots
- Rank the peptides using the Euclidean distance including results from 1D, 2D, 3D and 4D comparisons

UCSC Genome Browser Track Building, Sep. 2020-

- Create BED (Browser Extensible Data) files based on lncRNA database online
- Convert BED file to executable BB (BigBed) file using UNIX scripts
- Create folders and files following the protocol of building UCSC track hubs
- Add tracks on track hubs of UCSC based on the folders and files created for the visualization of the data

Exploration Of tRNAs As Predictors of Tumor Type, May 2020-

- Use machine learning algorithms such as K-means Cluster to cluster the tRNA data points
- Adopt ML feature importance methods (such as Logistic Regression and Random Forest) to compare various tRNAs' importance in different cancers
- Use pathway analysis to classify tRNAs in terms of upregulation and downregulation and score them in their differential expressions in tumors

• draw the corresponding heatmap using python package SEABORN

Next Generation Sequencing Analysis, May. 2020-
  • Conduct GSEA and GOrrila analysis to select genes of interest from the RNAseq gene expression data
  • Do regressions between gene counts and CFAB (Comprehensive Functional Assessment Battery) values to identify genes with differential abilities
  • Compare genes from young, older and elderly mice samples in terms of parameters from regressions
  • Visualize the data in volcano and heatmap plots

Multiple Approaches Toward Identification of Choline Degrading Genes via Whole-Metagenomic Sequencing, Aug. 2020-
  • Use ShortBRED and sequence-similarity networks to identify shared protein families from broader reference sequences
  • Conduct KMER based association analysis
  • Apply multiple sequence alignment to identify conserved residues of functional importance

Options Call Parameter Extraction, Sep. 2019- Dec. 2019
  • Build Scrapper to download 10Q & 8K from SEC websites for specified entity using python packages Request, BeautifulSoup and Selenium
  • Convert the PDF files into text using python library PyPDF3
  • Identify the values associated with the parameters of interest from the SEC text
  • Display the parameters in Tabular form

Suspended Stock Sentiment Analysis (Product Control of Credit Suisse), Sep. 2019-Dec. 2019
  • Scrape news articles related to stocks of interest from designated websites using python libraries such as Request, BeautifulSoup and Selenium
  • Preprocess the text data using NLP tools such as nltk, spacy and Gensim
  • Label the sentiment of news articles with the stock prices scraped by using python module Quandl
  • Apply different sentiment analysis models using various machine learning/deep learning algorithms to predict the sentiments of news articles pertinent to a given stock

Data Analysis and Predictive Model Building, May 2019-Dec. 2019
  • ETL (Extract, Transform, Load) Perinatal data (baby, mom, fetus data) from the Epic Electronic Medical Records
  • Integrate that data into a dataset using SQL and Python
  • Use Tableau to create dashboards to find data insights for patient care, patient safety, quality/process improvement
  • Use machine learning algorithms to make better predictions of the risk of PPH (Post-Partum Hemorrhage)

Optimal Workstation Design for Radiation Oncologists in UNC Cancer Hospital, May 2019- Dec. 2019
  • Do literature review of ergonomic workspace design
  • Analyze video and text data from experiment
  • Use machine learning algorithms to process data and validate the quantification of metrics involved in the ergonomic analysis such as cognitive workload, situation awareness and performance

Impact of Virtual Urgent Care on Time, Cost, and Patient Satisfaction, January 2019-May 2019
  • Evaluate virtual urgent care on four domains- access to care, effectiveness, experience, and economic impact
  • Compare the UNC virtual urgent care with in-person urgent care centers on these four domains, and develop broad framework for evaluation of virtual urgent care centers
  • Use Python libraries 'googlemaps', 'uszipcodes', 'pandas', and 'numpy' for feature engineering

Crowdsourcing and Image Analysis Project in UNC Kenan Flagler Business School, Jan. 2019- Aug. 2019
  Data Collection
  • Develop and implement code to scrape image data from Design Hill website, a specialized logo competition platform
  • Design local database for data storage and normalize the database using SQL skills
  • Execute website scrape, data cleansing, and data pre-processing for future analysis

Data Analysis

• Use OpenCv to quantify the similarity between images

• Use K Means Clustering to calculate the optimal number of clusters of images from each logo contest and submitted by each designer as an indicator of the diversity of the works from contests and designers

• Research new methodologies and machine learning algorithms that could produce better, more accurate, measures of similarity, originality, learning, etc.

Sentiment analysis of Yelp restaurant reviews, October, 2018

• Use python to scrape customers' review data from Yelp

• Sample and label the dataset, randomly split them into training and test sets

• Select feature representation using LightSIDE and Python coding

• Compare the accuracy of three classifiers, namely, Naïve Bayes, SVM, Logistic Regression and Random Forest

• Use the classifier with the best accuracy to predict the sentiment of customers and recommend restaurant based on the prediction

Information Gathering Plan for Reducing Pharmacy Inventory at Local Pharmacy May，2018

• Use system analysis models to analyze the flow of medicine in one Harris Teeter pharmacy

• Come up with a proposal as to augment the efficiency of pharmacy inventory operations and to reduce the workload of pharmacists

PROFESSIONAL EXPERIENCE

Research Associate: UNC-CH, School of Medicine, Chapel Hill, N.C., U.S.A. (Feb. 2021-present)

Research Specialist: UNC-CH, School of Medicine**,** Chapel Hill, N.C., U.S.A. (Mar. 2020 – Feb. 2021)

Data analyst of Bioinformatics & Analytics Research Collaborative (BARC)

Business Analyst (toward INLS 795 course of UNC-CH): Credit Suisse, Raleigh, N.C., U.S.A. (Oct.2019 – Dec. 2019)

Data Analyst Intern: UNC Center for Maternal and Infant Health (UNC CMIH), N.C., U.S.A. (May 2019 – Dec. 2019)

Research Assistant: UNC Department Of Radiation Oncology, UNC-CH, N.C., U.S.A. (May 2019 – Dec. 2019)

Carolina Applied Informatics Research Group Member: School of Nursing, UNC-CH, N.C., U.S.A. (Jan. 2019 – May 2019)

Graduate Research Assistant: Kenan Institute of Private Enterprise, Chapel Hill, N.C., U.S.A. (Aug. 2018 – Aug. 2019)

Writer and Translator: Craft Translation (New York) (May. 2013 – Oct. 2018)

Realize the glocalization of international brands including UPS, Gulfstream, Cisco, Dove and Mars

Lecturer: SIAS International University, China (U.S. Sponsored) (Aug. 2007 – Oct. 2011)

Lecturer: PLA University of Foreign Language, China (Aug. 2001 – Jul. 2007)

PUBLICATIONS

Paper: *Impact of Virtual Urgent Care on Time, Cost, and Patient Satisfaction: A Comparative Effectiveness Study,* published by *Journal of Patient Experience*
https://journals.sagepub.com/doi/full/10.1177/2374373520981487

Constitution, Attributes and Assessment of Mobile English Teaching, published on *Distance Education, China*