

Keshav Sompura

DATA SCIENTIST

Thousand Oaks, California

(805)-807-7002 | keshavsomp@gmail.com | github.com/ksompura | linkedin.com/in/keshav-sompura

Personal Profile

Experienced Data Scientist with a focus on NLP, statistical modeling, and machine learning. Currently working as a Research Data Scientist at Lund University, conducting large-scale social media analysis and researching AI's impact on society-wide thought patterns. Previous experience includes identifying policy defaults through building machine learning models at CA State Compensation Insurance Fund, successfully identifying revenue opportunities through NLP-based client segmentation as a data analyst intern, and completed a Master's thesis creating AI models for suicidal ideation detection on social media sites using BERT transformer models.

Experience

Research Data Scientist

Thousand Oaks, CA

Lund University

May 2023 - Present

- Collected and analyzed >2 million social media posts related to AI, machine learning, business intelligence, digitization, etc.
- Analyzed changes in social thought patterns over a 10 year period across 10 industry sectors and 6 topics of discussion.
- Analyzed the effects of chatGPT's release on >50 different language characteristics by building 3 different regression models to ensure robust results.
- **Technical Skills:** Python, LIWC, NLP, Text Analysis, Statistical Modeling, Panel Regression, Fixed-Effects Regression, Random- Effects Regression, Pooled OLS Regression, Topic Modeling, Clustering, Unstructured Data, Hypothesis Testing.

Data Scientist

Los Angeles, CA

State Compensation Insurance Fund

June 2023 - July 2023

- Worked with a corporate finance manager to build multiple machine learning models to identify which insurance policies will default.
- Identified significant and important variables that would lead to policy defaults.
- **Technical Skills:** R, SAS Miner, Machine Learning Modeling, Variable Importance, Structured Data.

Harnessing AI for Suicidal Ideation Detection - Master Thesis

Lund, Sweden

Lund University

March 2023 - May 2023

- Developed BERT transformer models to probabilistically flag social media messages for at risk suicidal individuals.
- Contacted and consulted with multiple domain experts during data collection, research, and data validation processes.
- Performed robust and rigorous hyperparameter tuning and testing to ensure accuracy and precision of model results.
- **Technical Skills:** Python, pytorch, NLP, Text Classification, Transformers, BERT, Machine Learning, Deep Learning, Pattern Discovery, Social Media, Unstructured Data.

Data Analyst - Intern

Agoura Hills, CA

Growing Outdoors

September 2021 - April 2022

- Discovered distinct client segments with NLP methods which identified opportunities to increase revenue.
- Designed, developed, and deployed interactive dashboards for 6 different facilities.
- Streamlined data cleaning and transformation time by 60% through Excel macros.
- Collaborated with two company directors to create a clear goal using their previously compiled data to classify client behaviors and geographical markets.
- **Technical Skills:** NLP, Topic Modeling, Text Data, Dashboards, Data Visualization, R, tidyverse, ggplot2, shiny, Excel.

Bus Captain, Counselor, Head Lifeguard

Agoura Hills, CA

Camp Kinneret Summer Day Camp

June 2017 - August 2022

- Supervised and led a team of 5 staff and 4 interns on a bus daily to ensure a safe and fun environment for 50+ campers.
- Led two groups of 15 campers, building an environment of fun, creativity, and encouragement.
- Trained, led, and supported a lifeguard staff of 6 employees.
- Oversaw and organized the aquatics program of 60 students per hour and up to 300 students per day.

Projects

CNNs and Vision Transformers for Facial Emotion Recognition

Lund, Sweden

Lund University

March 2023

- Created deep learning models for recognizing and probabilistically classifying people's facial emotions from images.
- Optimized hyper-parameters using random search to improve model performance.
- Built and tested numerous convolutional neural networks and fine-tuned vision transformers.
- **Technical Skills:** Python, CNNs, Vision Transformer, tensorflow, pytorch, Computer Vision.

Strategic Cost-Saving Models for AOB Travel Company

Lund, Sweden

Lund University

February 2023

- Analyzed position in the industry and found an area to develop a cost leadership strategy.
- Processed large JSON datasets to process into usable data format.
- Built, tested, and optimized ML models using company data to establish a cost saving measure that would lead to a 5-10% reduction of extraneous costs.
- **Technical Skills:** R, Machine Learning Modeling, Data Cleaning and Mining, tidymodels, tidyverse, biglasso.

NLP Second Hand Clothing Industry Analysis

Lund, Sweden

LAMBDA - Lund Data Science Society

January 2023 - February 2021

- Scraped and analyzed +200k google play app reviews for 6 second-hand clothing companies and performed sentiment analysis to track customer sentiment over time.
- Clustered reviews into topics in order to identify well-liked practices and critical issues so they can be implemented or flagged.
- **Technical Skills:** Python, APIs, NLP, Topic Modeling, Data Wrangling, Sentiment Analysis, pandas, matplotlib, seaborn.

Market Analysis: Bellabeat Case Study

Thousand Oaks, CA, US

Google - Data Analytics Professional Certification

December 2021

- Joined and analyzed >1 Million row Fitbit datasets using statistical analysis to discover intra-day and week customer behavior patterns.
- Created informative visualizations, an actionable report for the company to improve an existing product, and proposed a data-driven path to cultivate brand loyalty.
- **Technical Skills:** Market Analysis, Big Data, R, Data Visualization, tidyverse, Interactive Graphs.

Education

Lund University

Lund, Sweden

MSc in Data Analytics and Business Economics

August 2022 - June 2023

- Learned a wide array of machine learning model types and when to apply each type with included regularizations to counter bias and noise inherent to data problems.
- Built ML models and created data-driven strategies for IKEA, AOB Travel, and H&M managers.
- Co-Founded LAMBDA - Lund Data Science Society.
- **Courses:** Advanced Machine Learning, Analytics-based Strategic Management, Data Visualization, Machine Learning from a Regression Perspective, Databases, Deep Learning and Artificial Intelligence Methods.

University of California, San Diego

San Diego, CA, US

BA in Economics - **Focus:** Econometrics and statistical modeling

September 2017 - June 2021

- GPA: 3.7 / 4.0 - Graduated with Department Honors.
- Secretary of the swim club and competitive swimmer.
- Active member at International House - provided support and became friends with many international students.

Sciences Po

Paris, France

Exchange - International Business and Economics

August 2019 - December 2019

- Education and cultural exchange.
- Became conversational in french.

Skills

Programming Python (Pandas, PyTorch, NumPy, Scikit-learn, Tensorflow, Pytorch, Matplotlib), R (tidyverse, glmnet, caret, ggplot2, shiny), BigQuery, Jupyter, SQL, SAS Miner, Hypothesis Testing

ML & DL Gaussian Mixture Models, Double Machine Learning, Causal Trees & Forests, Bayesian Models, K-Means Clustering, Elastic Net, PCA, Random Forests, XGBoost, Multivariate Regression, Linear & Logistic Regression

Miscellaneous Math, Statistics, Git, Github, Azure Data Studio, Tableau, Microsoft & Google Suite, Dashboards, Power BI, Slack, Overleaf, R Markdown, Python Notebooks

Languages

English Native proficiency

French Conversational proficiency

Swedish Elementary proficiency

Gujarati & Hindi Semi-proficient in vocal understanding, but limited speaking ability