SATEJ SOMAN

satejsoman.com

SKILLS

COMPETENCIES: data pipelining, distributed systems, stochastic simulation, geospatial visualization, time series analysis, econometrics, policy evaluation, Bayesian inference, machine learning
PROGRAMMING LANGUAGES: Python, R, Java, Haskell, C/C++, Go, Stata, Mathematica, MATLAB, JavaScript
SYSTEMS & DATABASES: Spark, Kubernetes, ElasticSearch, Cassandra, Postgres/PostGIS
LANGUAGES: Spanish (fluent), Marathi (fluent), French (conversational), Hindi (conversational), German (basic)

EXPERIENCE

COVID International Working Group / Head of Data Science

JUN 2020 - PRESENT, CHICAGO, IL

Advised national and sub-national governments in India and Indonesia on COVID-19 lockdowns, economic recovery, and vaccine distribution policies.
Developed novel, computationally-efficient Bayesian estimation models for key epidemiological quantities.
Led team of data scientists to build a cloud-hosted data pipeline to deliver daily-updated insights to policymakers (examples available at adaptivecontrol.org).

Mansueto Institute for Urban Innovation / Research Engineer

JAN 2019 - JUN 2020, CHICAGO, IL

Led development and data analysis for the Million Neighborhoods Project (millionneighborhoods.org), the first global map identifying slums through topological analysis.
Created new algorithms for extracting street block and cadastral parcel geometries from open-source geospatial data.
Responsible for scaling data analysis and processing to terabyte-scale and for onboarding new researchers.

Palantir Technologies / Forward-Deployed Engineering Lead

AUG 2014 - AUG 2018, PALO ALTO, CA + SINGAPORE, SG + PARIS, FR + LONDON, UK

Responsible for developing and implementing novel money-laundering and cyber intrusion detection methods used by banks and insurance firms to prevent financial crimes and comply with international regulations.
Drove development of a geospatial risk analysis tool used to negotiate reinsurance pricing in Latin America & Europe.
Led R&D group building system for statistical inference of relational keys across database systems.
Redesigned engineering interview and hiring process for Singapore, Paris, and London offices.

Flutter (acquired by Google) / Software Engineer

MAY 2012 - MAY 2013, MENLO PARK, CA

Developed toolset to measure computer vision algorithm performance according to established machine-learning metrics, including ROC characteristics and precision-vs-recall curves.
Deployed regression testing framework on Hadoop to reduce product testing time by 10x.
Experimented with data-compression algorithms to reduce support vector machine training time.

EDUCATION

University of Chicago / Harris School of Public Policy

OCT 2018 - MAY 2020, CHICAGO, IL

M.S. Computational Analysis & Public Policy (certificate in Financial Analysis & Policy)

University of California, Berkeley / College of Engineering

SEP 2010 - MAY 2014, BERKELEY, CA

B.S. Electrical Engineering & Computer Science, Materials Science & Engineering (minor in Physics)

PUBLICATIONS

Real-Time Epidemiological Prediction and Control
(preprint)
Bettencourt, L. & **Soman, S.** (2020).
Mansueto Institute for Urban Innovation Research Paper, (24).

Worldwide Detection of Informal Settlements via Topological Analysis of Crowdsourced Digital Maps
(peer-reviewed, accepted for publication)
Soman, S., Beukes, A., Nederhood, C., Marchio, N., & Bettencourt, L. (2020).
ISPRS Int. J. Geo-Inf., *9*(10).
preprint: preprints.org

Adaptive Control of COVID-19 Outbreaks in India: Local, Gradual, and Trigger-based Exit Paths from Lockdown
(working paper)
Malani, A., **Soman, S.**, Asher, S., Novosad, P., Imbert, C., Tandel, V., ... & Shen, D. (2020).
NBER Working Paper, (No. w27532).
preprint: nber.org/papers

PATENTS

Data Item Aggregate Probability Analysis System
#US10691756B2; lead inventor
System to allow selection and filtering of datasets by geospatial index and by configurable risk score.

Systems and Methods for Annotating Datasets
#US20190179936A1
Application of the Flajolet-Martin algorithm to identify potential database joins in data warehousing systems.