SUBHAJIT DAS

Data Modeling + Data Engineering | Machine Learning | User Research
W: **https://subhdas.github.io/** | **LinkedIn** **E** : **das.connect@outlook.com**
M : +1 404 543 1580

Over 5 years of experience in machine learning model development with proven success (research and industry) in data modeling, visualizations, algorithm design, and interactive predictive analytics. Skilled to build classification, regression, clustering, topic modeling, web-scraping, deep learning for image classification and sequence modeling etc. Conducted research at Visual Analytics Lab at Georgia Tech and with DARPA on AI / Machine Learning model selection, scalable data visualizations, web frameworks, and cloud computing to include user interaction data in modeling.

_EXPERTISE

Tools: Python, PostgreSQL, SQL, AWS, MongoDB, Spark, Linux, Scala, Javascript, Typescript
Packages: Scikit-Learn, NumPy, SciPy, Pandas, NLTK, BeautifulSoup, Matplotlib, Jupyter Notebook, D3
Machine Learning: Statistical Analysis, Linear/Logistic Regression, Classification, Clustering, Graph Theory, Regularisation, Deep learning in Pytorch and Tensorflow (CNN, RNN), Hypothesis testing

---

_ACADEMICS

GEORGIA INSTITUTE OF TECHNOLOGY, College of Computing, Atlanta, USA
PhD in Computer Science - Interactive Machine Learning + Visual Analytics   **2014 - 2020**
Advisor : Dr. Alex Endert
Working with domain experts, have developed novel interaction techniques (for the web) with multiple machine learning models using model steering, and data engineering for optimizations driven by user interactions. Have deployed custom-designed models and algorithms to adjust modeling tasks to domain specific data problems.
Interests : Machine learning, Data Modeling, Data Visualization, User Research, HCI + AI, Visual Analytics

MS in Computer Science - Visual Analytics
- Machine Learning, Visual Analytics, AI, Graphics and Optimization, Experimental Design, Computer Vision, HCI
- Data Science Certificate from John Hopkins University (Coursera.org)

UNIVERSITY OF PENNSYLVANIA, PENNDESIGN & WHARTON, Philadelphia, USA
M. Arch. Design Computing (Evolutionary Algorithms, Digital Fabrication)   **2008 - 2009**
Swarm Intelligence, Evolutionary Computing, Multi Agent Systems, Strategy Management (PgDBA)

JADAVPUR UNIVERSITY, Kolkata, India
B. Arch Engg. 1st Class | Leed AP Healthcare Design Based on Human Cognition   **2003 - 2008**

_WORK EXPERIENCE

3M Healthcare (HIS) - NATURAL LANGUAGE UNDERSTANDING, Atlanta   **Mar 2021 - Current**
Data Scientist | Research Engineer
Under the Natural Language Understanding (NLU) team, motivated to solve real-world healthcare problems using AI/ML modeling techniques in natural language processing (NLP). Developing API's for novel metrics and statistics to improve 3M's core NLP engine by analysing user feedback streams from an AI expert system.

VISUAL ANALYTICS LAB + DARPA, Atlanta   Aug 2016 - Jan 2021

Researcher under Dr Alex Endert | Javascript, React, Vue, D3, Python, Tensorflow

PhD Work: Researching novel interactions for web interfaces to allow curation, optimization, selection and steering of multiple machine learning models to broaden access to ML. Working with domain experts, developed methods to interactively create novel data models to navigate model space; optimize and select models using techniques such as multi-model steering, and interactive objective functions for analytics based processes.

DARPA funded projects: Collaborated with Tufts University, and University of Wisconsin to prototype ML applications to enable interactive data augmentation and model selection to democratize access to AI solutions.

UPS INDUSTRIAL ENGINEERING, Atlanta   Jun - Aug 2020

Visual Analytics Intern | Typescript, Angular, Leaflet

Developed a geo-visualization system and a data model to make sense of transportation and operation teams' optimization models. Developed an Angular application with Leaflet JS as the geo-visualization technology.

BOSCH RESEARCH, Sunnyvale   May - Aug 2019

Visual Analytics Intern | Python, Pytorch, TorchText (Patent filled)

An explainable AI application using matrix optimizations on multiple data formats. Crowd-sourced user feedback from Amazon's Mechanical Turk. Wrote a novel neural network architecture using Pytorch, and NLTK. Tested the novel neural architecture design on Amazons' product review dataset to explain prediction of sentiments.

AUTODESK, Boston   May - Aug 2018

Software Engineering Intern | Javascript, C++

Wrote microservices to unify data and parameters across multiple desktop applications based on user feedback. Worked with legacy code written in C++ to integrate microservices written in javascript and Google Sheets API.

MICROSOFT RESEARCH, Cambridge, UK   May - Aug 2017

Research Intern | Typescript, Node, React

Prototyped STORYBOARD - a web-based slideshow builder, facilitating users to focus on content development by automating slide organization and management. Wrote a React app from scratch using Typescript and NodeJS.

AUTODESK, Atlanta   Jan - Aug 2016

Software Development Intern | C#, .net

Prototyped a plugin written in C# generating a large number of automated 3D floor plans helping healthcare planners pick a suitable option. Conducted user study with experts from Perkins Will to further validate the tool.

AUTODESK, Boston   Jun - Aug 2015

Software Development Intern | Python

Worked on the development of a web-based 3D design platform in ThreeJS. Incorporating domain expertise prototyped an extension written in Python enabling automated delivery of 3D model options for facility planning.

MACHINE LEARNING APPLICATIONS | **Researcher | Python, Pytorch, Tensorflow, Javascript**

- **LEGION: Visually compare modeling techniques for regression**   2020
  Designed and deployed a web application that assists users in deciding how to construct regression models: (1) By hyperparameter tuning, or (2) By Feature engineering. Motivated to help data analysts compare models to learn tradeoffs in various modeling techniques. At Visualization in Data Science, IEEEVIS 2020
- **INMACS: Interactive modeling and comparison of sentiments from sequence data**   2020
  Developed an interactive AutoML system that helps users to iteratively preprocess data (using natural language processing) to build topic and sentiment models. DaSH Workshop at KDD Conference 2020.
- **QUESTO: Interactive construction of objective functions**   2019

Built a system that allows users to interactively design objective functions to model multiple classifiers using an AutoML optimizer. Presented at EuroVis Conference 2020.

- **GEONO-CLUSTER: Interactive Visual Cluster Analysis for Biologists, Georgia Tech**   2018 - 2019
Domain study with biology researchers, to build an interactive clustering system that incorporates user feedback in model recommendations for data exploration, IEEE TVCG 2020.
- **AUTODESK Machine Learning Group | Python**   **Nov - Aug 2016**
Developed models to help instantaneously predict energy simulation results without the need to endlessly wait for model runs. Compared performance accuracy with Deep Neural Networks and Boosted trees.

_PATENTS

With BOSCH RESEARCH, SUNNYVALE, CA: (filled, under processing)
Improving Deep Neural Networks via Prototype Factorization, Zeng Dai, Panpan Xu, Liu Ren, Subhajit Das

----

_SKILLS

PROGRAMMING: **Proficient**: Typescript, Javascript, C#, Java, Python, Tensorflow, Pytorch, Keras, Pyspark
Working knowledge: C++, Rust, Scala, C, Web Assembly, Android, Arduino, OpenCV
WEB: NodeJS, React, Angular, VueJS, D3, Vegalite, ThreeJS, Full-stack
AWS, MongoDB, SQL, PostgreSQL Google Cloud Platform, Docker, AWS Sagemaker
USER RESEARCH: Usability tests, Mixed-method research, Crowd-sourced surveys, Hypothesis Testing
OTHERS: Hadoop, Unreal Engine, Unity 3D, Processing, Tableau, Power BI

_PUBLICATIONS

JOURNAL/CONFERENCE PAPERS

1. Florina Dutt, **Subhajit Das**, **Fine-grained Geolocation Prediction of Tweets with Human Machine Collaboration,** Arxiv Preprint - https://arxiv.org/pdf/2106.13411.pdf.

2. **Subhajit Das**, Alex Endert. **CACTUS: Detecting and Resolving Conflicts in Objective Functions**. Arxiv Preprint - https://arxiv.org/pdf/2103.07805.pdf.

3. Dylan Cashman**,** Shenyu Xu, **Subhajit Das**, Florian Heimerl, Cong Liu, Shah Rukh Humayoun, Michael Gleicher, Alex Endert, Remco Chang. **CAVA: A Visual Analytics System for Exploratory Columnar Data Augmentation Using Knowledge Graphs** IEEE Transactions on Visualization and Computer Graphics 2021. Accepted in IEEE VAST Conference 2020.

4. **Subhajit Das**, Shenyu Xu, Michael Gleicher, Remco Chang, Alex Endert. **QUESTO: Interactive Construction of Objective Functions for Visual Analytics**. Computer Graphics Forum, 39: 153-165. In EuroVis Conference 2020.

5. **Subhajit Das,** Dylan Cashman, Remco Chang, and Alex Endert. **Gaggle: Visual analytics for model space navigation.** In Graphics Interface, 2020.

6. **Subhajit Das,** Bahador Saket, Bum Chul Kwon, Alex Endert. **Geono-Cluster: Interactive Visual Cluster Analysis for Biologists.** IEEE-Transactions on Visualization and Graphics 2020.

7. **Subhajit Das,** Dylan Cashman, Remco Chang, and Alex Endert. **BEAMES: Interactive multi-model steering, selection, and inspection for regression tasks.** IEEE computer graphics and applications, 39(5):20–32, 2019.

8. Dylan Cashman, Shah Rukh Humayoun, Florian Heimerl, Kendall Park, **Subhajit Das,** John Thompson, Bahador Saket, Abigail Mosca, John Stasko, Alex Endert, Michael Gleicher, and Remco Chang. **A user-based visual analytics workflow for exploratory model analysis**. Computer Graphics Forum, 38(3):185–199, 2019. In EuroVis Conference 2019.

9. Po-Ming Law, **Subhajit Das,** and Rahul C. Basole. **Comparing Apples and Oranges: Taxonomy and Design of Pairwise Comparisons within Tabular Data**. ACM Conference on Human Factors in Computing Systems (CHI) 2019.

10. Mathew Shwartz, **Subhajit Das. Interpreting non-flat surfaces for walkability analysis.** Society for Computer Simulation International and SIMAUD, 2019.

11. Emily Wall, **Subhajit Das**, Ravish Chawla, Alex Endert. **Podium: Ranking Data Using Mixed-Initiative Visual Analytics.** IEEE Transactions on Visualization and Computer Graphics 2018, In IEEE VAST Conference 2017.

12. Mohammad Asl Rahmani, **Subhajit Das,** Barry Tsai, Ian Molloy, and Anthony Hauck. **Energy Model Machine (EMM) - Instant Building Energy Prediction using Machine Learning**. In eCAADe Conference 2017.

13. **Subhajit Das**, Colin Day, John Haymaker, Anthony Hauck, Diana Davis. **Space Plan Generator: Rapid generation & evaluation of floor plan design options to inform decision-making**. In Acadia Conference 2016.

WORKSHOP PAPERS

14. **Subhajit Das,** Alex Endert**. CACTUS: Detecting and Resolving Conflicts in Objective Functions.** Visual Data Science (VDS) Workshop, KDD and IEEE-Vis 2021

15. **Subhajit Das,** Panpan Xu, Zeng Dai, Alex Endert, Liu Ren, **Interpreting Deep Neural Networks through Prototype Factorization**, Deep Learning and Clustering, DLC Workshop, at IEEE ICDM 2020.

16. **Subhajit Das,** Alex Endert. **LEGION: Visually compare modeling techniques for regression.** In IEEEVIS, Symposium on Visualization in Data Science (VDS) 2020.

17. **Subhajit Das,** Florina Dutt. **InMacs: Interactive modeling and comparison of sentiments from sequence data**. Workshop on Data Science with Human in the Loop (DaSH). In KDD Conference 2020.

18. **Subhajit Das,** Dylan Cashman, Remco Chang, and Alex Endert. **BEAMES: Interactive multi-model steering and inspection for regression tasks**. In IEEEVIS Symposium on Visualization in Data Science (VDS), 2018. [**Best Paper Award].**

19. **Subhajit Das,** Andrea McCarter, Joe Minieri, Nandita Damaraju, Sriram Padmanabhan, D.Horng (Polo) Chau. **ISPARK: Interactive Visual Analytics for Fire Incidents and Station Placement**. Workshop on Interactive Data Exploration and Analytics (IDEA) at KDD conference 2015.

PRESENTATIONS/TALK/WORKSHOPS

20. **Subhajit Das, Enabling access to machine learning to power users**, Shandong University, China, Jan 2020.

21. **Subhajit Das, Confluence : Machines + Human expert in Design Automation,** NJIT, USA, Mar 2018.

22. **Subhajit Das,** Arifin Mohiuddin, **Conducted workshop on Design optimization & exploration with Autodesk,** Sweden, 2018.

23. **Subhajit Das,** Florina Dutt, **Job Location Mismatch in Atlanta, Based On Available Workforce from Big Data,** Esri User Conference, San Diego, USA, June 2016.

24. **Subhajit Das,** John Haymaker, Anthony Hauck, **Conducted workshop on Space plan generation with Autodesk and Healthcare planners from Perkins Will,** Boston, USA, Aug 2016.

25. **Subhajit Das, Data model and processes for building programming,** Digital Building Lab Symposium, Georgia Institute of Technology, Atlanta, USA, Aug 2015.

POSTERS

26. Emily Wall**, Subhajit Das,** R. Chawla, and A. Endert, **Podium: Ranking Data Using Mixed-Initiative Visual Analytics**, CRA-W Grad Cohort, San Diego, CA, April 2016.

27. Nandita. Damaraju, **Subhajit Das,** Andrea McCarter, Joe Minieri, Sriram Padmanabhan, D. H.(Polo) Chau, **Mitigating fire risks using visual and data analysis,** Bloomberg: Data for Good Exchange, New York, USA, Sep 2015.

28. **Subhajit Das,** Florina Dutt, **Design ornamentation & fabrication by multi agent systems,** SIGGRAPH Posters 2012.

_COURSES

MACHINE LEARNING:	Machine learning, ML for Finance, Computer Vision, Visual Analytics, AI, Deep Learning (Stanford)
HCI:	Principles of HCI, User Interfaces, Prototyping Interactive Systems, Information Visualization
GRAPHICS:	Animation, Computer Graphics, Simulation of Complex Systems, Human Robot Interaction
COMPUTING:	Computability and Algorithms, Modeling and Simulation, Sonification Lab, Spatial Optimization

---

_REVIEWER

IEEE VAST, Infovis, CHI, Computer and Graphics, IEEE TVCG, Eurovis, Computer Graphics Journal, Graphics Interface, TAD Journal