Vindhya Singh — 25 March 1994 — Find me here

Email : vindhyakidak@gmail.com
Mobile : +49 1523 6425768
Referees: Prof.Dr.Bjoern H. Menze (bjoern.menze@uzh.ch) and Prof.Dr.Georg Groh (grohg@in.tum.de)

Github Profile
LinkedIn Profile

SKILL SET

Computer Languages	Python, R
ML Tools and Products	PyCharm, AWS Sage Maker, Google Colab, Jupyter
ML/DL Libraries and Frameworks	PyTorch, Tensorflow, Pandas, NumPy, Scikit-Learn, etc.
Graph Based Tools	Gephi, KH Coder
Augmented Reality Tools	Unity 3D, Vuforia
DevOps	Docker, Kubernetes, Jira, Ansible
BI and ETL Tools	Tableau, Quicksight, QlikView, Informatica, PowerBI
Operating Systems	Ubuntu, Windows, macOS
Databases	Oracle, MySQL, Amazon Redshift
Others	Hadoop, Spark, Docker, Kubernetes, Openshift
Spoken Languages	English (C2), German (A2), Hindi (Native)

EXPERIENCE

IMPRS-IS and University of Stuttgart — Sep 2023 - Present
PhD Candidate — *Stuttgart, Germany*

· Working on NLP methods in the application area of Social Science and Data Visualisation.

iABG — Dec 2020 - Aug 2023
Data Scientist — *Munich, Germany*

· Working with the German Armed Forces on projects in the domain of computer vision and NLP based on predictive decision modeling. One of the projects can be found at SPARTA website.

Technical University of Munich — June 2020 - Nov 2020
Student Research Assistant - Augmented Reality — *Munich, Germany*

· Developing Augmented Reality (AR) Application to allow interactive learning for students by visualizing chemical reactions (Unity and Vuforia).

Amazon — June 2018 - May 2020
Working Student - Data Engineer — EU Fashion — *Munich, Germany*

· Owned the In-Season Apparel and Shoes ASINs buys project. Increased the Contribution Profit by a million euros in a quarter through quantitative analysis of data trends. It assisted various vendor and product managers on a bi-weekly basis in managing their product inventory.
· Designed interactive dashboards on Excel (through VBA and Power Query) and Tableau (via MySQL database connections) for multiple projects. Used PowerShell and Python scripting for automatic updates and report generation.
· Facilitated product managers across various teams to optimize the Buy Box percentage by formulating metrics and reports, and analyzing the input data that led to data-driven decision-making.
· Implemented Data Processing pipeline from data fetch (through SQL queries) to data visualization (Tableau/Quicksight/MS-Excel). Performed forecast modeling using algorithms such as Random Forest (R and AWS Sagemaker).

StudySmarter March 2018 - August 2018
Member of IDP Project *Munich, Germany*

· Implemented Machine Learning Algorithms for automatically generating tags and classifying documents into respective subjects for a corpus (Python/Tensorflow).
· Created a comprehensive pipeline for document tag generation and classification after a thorough evaluation of various supervised and unsupervised machine learning algorithms (SQL/Excel/Python).

Cognizant Technology Solutions November 2015 - July 2017
BI Engineer *Pune, India*

· Performed a novel method of analyzing personal and behavioral traits of clients through the use of AI and visualizing the findings on QlikView in an interactive dashboard.
· Discovered exclusive areas for gaining new customers and increasing the profit margin of the company by analyzing the trends of stakeholders in the life-sciences domain of Cognizant.

EDUCATION AND SUMMER SCHOOLS

M.Sc (Informatics, Major: Deep Learning) — Technische Universität München (Germany)
October 2017 - Sep 2020

Eastern European Machine Learning (EEML) Summer School (Virtual) — Krakow, Poland
July 2020

ATHENS Programme, Course : From Complexity to Intelligence — Telecom ParisTech (France)
March 2019

Bachelor of Technology (Computer Science) — Galgotias University (India)
August 2011 - June 2015

IABG PROJECTS AND SEMINAR

SPARTA — 2022
Python, PyTorch, KH Coder, Gephi, Relatio, Natural Language Processing, Deep Learning
SPARTA Project Website
The project name SPARTA stands for Society, Politics, and Risk with Twitter Analysis. The interdisciplinary research project SPARTA done in collaboration with the ***Universität der Bundeswehr München brings together political science and computer science and builds a state-of-the-art infrastructure for NLP analyses of real-time and historical Twitter data.***

Worked on 2 major use cases:

- *Elections Use Case:* ***Sentiment Analysis and Stance Detection Analysis*** *on real-time Twitter data for the Bundestag Elections of 2021. Results for the same can be found here. In Spring 2022, worked on evaluating the election campaign in NRW in real-time. Results for the same can be found here. Also worked on* ***analyzing the linguistic capabilities of deep learning models****, mainly BERT-based models. For both these use cases, the tweets were in German and English.*
- *Riots Use Case: Using few-shot learning, worked on* ***identifying and classifying the user communities for the Capitol Hill Riots of 2021****. Evaluated and analyzed the* ***language used in the Twitter discourse of the Capitol Hill Riots*** *using historical data from Twitter. Also worked on* ***hate speech detection, identifying the toxicity of the language, and identifying the language traits using the NRC Lexicon****. Worked on analyzing how Parts of Speech in the language are connected with each other with a significant focus on NOUN, VERB, ADV, that is,* ***N-ADJ (Modularity-based graph) N-ADJ (Degree centrality-based graph) N-Verb (Modularity-based graph) N-Verb (Degree centrality based graph). Three of our papers are upcoming for this use case by March 2023****.*

- *Additional Tasks: 1.) Co-Supervisor of a working student's Master's Thesis in the lab of **Dr. Alexander M. Fraser (LMU, Munich) on the topic of Multilingual Fake News Detection on Twitter Data using Deep Neural Networks.** 2.) Audio and video data analysis.*

Alfarabi — 2022
Python, PyTorch , SQLAlchemy, Natural Language Processing
M4P Project Website
*A custom-tailored, **multi-language approach to Fake News Detection and Analysis** for Lebanon and Afghanistan. Implemented various Natural Language Processing steps such as text pre-processing and language modeling for **Arabic, Dari, English, and Pashto.** Analyzed and evaluated the impact and role of Fake News Detection on the **Twitter discourse for Lebanon and Afghanistan** using deep learning-based NLP methods. For Lebanon, the analysis is centered around the Beirut port explosion while for Afghanistan, the central event is the fall of Kabul.*

ONEO — 2021
Python, PyTorch, Natural Language Processing, Elastic Search
***Poster presentation of our project at the Living Planet Symposium**, held in Bonn in May 2022. The project centered around Earth Observation (EO) and Media Analysis. The automatic geocoding of **social media information relates it to a geo-location.** EO imagery can be applied afterward to verify and evaluate the extracted information. Subsequent geospatial analyses enable area-wide investigations and improve the situational picture. Implemented comprehensive use of legacy remote sensing techniques and state-of-the-art Deep Learning approaches. Additionally, **event localization, Named Entity Recognition, and information retrieval** from online news / social media using Natural Language Processing.*

Scrabble — 2021
Python, Newspaper3k, news-please, Elasticsearch
*This project aimed to collect our knowledge on crawling and scraping news websites in one convenient and easy-to-use Python package. It is **database-agnostic** and can be used with a different queue or publication-subscription services. For this project, evaluated various **machine learning and neural topic modeling algorithms and approaches.***

More Robust Neural Networks — 2021
Python, Pytorch, Ray tune, R, Data Version Control (DVC)
*The project aimed to investigate the methods to increase the robustness of the neural networks. Implemented automated hyper-parameter tuning and data versioning control, leveraging the ability of the **meta-learning approach to tolerate degradations in a more fundamental way.** This may allow for reaching a higher level of robustness than in the case of traditional robustness training, therefore, making it harder to find adversarial examples. The results are evaluated on the MNIST dataset.*

FeSoLe: Few Shot Learning — 2021
Python, Pytorch, Tensorflow
*The project focussed on improving the efficiency of projects using the few-shot learning method. It also provided for the scope and evaluation of methods and technologies which cannot be performed within the application projects. This was a more exploratory research-oriented and less application-oriented project. Worked primarily with **BERT-based models and explored the cross-domain adaptation capabilities using few-shot learning.***

PAPAYA — 2020-2021
Python, Flask, SQLAlchemy, Dash, Docker, Software Engineering
PAPAYA is a Personal and Project Capacity Management Tool. Worked on creating an intuitive, convenient, and interactive dashboard that gives insights into the workload and available capacity of each team member in the department. For this purpose, the above-mentioned tools and technologies were used. It provides an interface to plan hours for projects and distribute capacities for individuals. Also, it enabled project leaders to monitor project status by comparing planning hours with actual hours worked on specific projects.

ACADEMIC PROJECTS AND SEMINAR

Exploring NLP Methods for diagnosis prediction from medical reports — 2020
Developed a novel approach utilizing NLP methods and models such as Topic Modelling, GloVe, and BERT to predict the diagnosis from a given radiology report for Master's Thesis.

Multilingual Theme Prediction for Amazon Customer Reviews — 2019
Implemented Siamese Network on Amazon Product Review dataset to generate similarity between two reviews written in different languages based on Product Category. Used LASER for multilingual embeddings. Zero-shot Transfer Learning from reviews trained in English to reviews written in German.

Guided Research : Recipe Substitute Recommendation — 2019
Food similarities used to generate recommendations for finding food substitutes in the recipes with an LSTM model with Center Loss trained on the Recipe 1M+ dataset.

Seminar: Deep Learning in Physics — 2019
Hidden Physics Models: Machine Learning of Nonlinear Partial Differential Equations

Distributed Data Mining — 2018
Explored and compared the performance of Hadoop and Spark. Performed Protein Sequence Prediction using Java, Python, Scala thereafter comparing their performance.

AWARDS AND CERTIFICATIONS

Awards and Recognition at IABG : *Poster presentation of ONEO at the Living Planet Symposium, held in Bonn in May 2022.*
Awards and Recognition at TUM : *MINGA Mentor, International Day (1st position), poster presentation in EEML 2020*
Awards at Amazon : *Research and Sustainability Volunteer, Amazon Alexa Beta Tester*
Awards at Cognizant : *Best-rated employee and 'High Flyer- Rising Star' award*
Awards during Bachelor's Degree : *Graduated with the highest percentage of marks in the Department of Computer Science and Engineering | Research Internship at IIT,BHU, India | Campus Ambassador for Accenture | Head Coordinator of the Literary Club*

Certifications
DataCamp: R Programmer (2018) | Machine Learning with Tree-Based Models in Python (2020)
AWS: Developing Machine Learning Applications (2019) | Machine Learning on AWS (2019)
Coursera: Mathematics for Machine Learning: Multivariate Calculus (2020) | Architecting with Google Kubernetes Engine: Foundations (2021)
Data Camp: Machine Learning with Tree-based models in Python (2020) | R Programmer Track (2018)
Hertie Data School: Data Science Summer School (2021)