

Deepak Singh

Assoc. Manager Data Science, Optum Global Solutions, Noida (Delhi -NCR)

Email: 09deepak010@gmail.com

Mob: [+91 - 7376202567](tel:+91-7376202567), [+91 - 9140690140](tel:+91-9140690140)



Academics

2015-2017		M.Sc. in Mathematical Statistics Indian Institute of Technology, Kanpur, UP	8.10 CGPA
2012-2015		B.Sc. in Statistics Banaras Hindu University (BHU), Varanasi, U.P. Gold Medalist	9.35 CGPA
2009-2011		Intermediate , U.P. Board C R T S V M Intermediate College, Maharajganj, U.P.	86.20 %
2007-2009		High School , U.P. Board A R D Inter College, Deoria, U.P.	80.17 %

Assoc. Manager Data Science & Inventor

Optum Global Solutions Jun'17 – Ongoing

Inventions

- Successfully filed two patent applications, both falling under the jurisdiction of the US Patent Office. Another patent is currently undergoing discussions with the Patent Review Board (PRB), with a scheduled meeting for further discussions.

Proactive Identification of Physician Burnout

Jan'21 – Ongoing

- Established **proactive physician burnout identification** and **intervention framework** within Databricks by leveraging EHR task history and demographic data.
- Created Python classes and definitions that are adaptable to various projects and designed a **Dashboard using Plotly Dash**. This dashboard displays the analytical insights and metrics crafted by the DS team, allowing for initial client feedback before passing it on to the UI designer for incorporation into the user interface.
- Formulated a **comprehensive Python definition** capable of managing **all facets of predictive modeling** within a singular framework.
- Integrated **Pydantic**-based data model, **pandas schema**, and **Great Expectations** for automated data quality assurance in the product.
- Facilitated a team consensus to adopt **Github Enterprise** as the central repository for codebase management. Promoted the utilization of Github for issue tracking and project management within the team. Enforced coding standardization by implementing **Black** formatting and **isort** for Python code linting. Utilized **SQLFluff** to ensure proper formatting of .sql files. Devised and executed a branch protection strategy for both development and production branches.
- Designed automated Python module for **raw data processing**, **quality checks**, **Snowflake upload**, and Github logging. Configurable for diverse file types and tables.
- Leveraged **MLFlow** for model registration and management in DataBricks. Utilized **Poetry** for uniform Python environment management across developers.
- Designed and executed strategy to manage and update production and development tables in snowflake using python automated scripts.
- Applied **Logistic regression with Log-F penalty** to handle the small dataset problems.
- Delivered the **key drivers of Burnout** at Individual physician level as well as organization level to enhance intervention effectiveness.
- As DS team SPOC, **contributed to UI portal** by conceptualizing, designing, and implementing key metrics.
- Took a key role in collaborating with the Machine Learning Review Board (**MLRB**) to secure approval for developed models aligned with Responsible Use of AI (**RUAI**) principles.
- Optimized execution time by writing time-consuming Python code into CPython

New Skills under Development

- Currently participating in a 5-month-long **Emerging Leadership Program (ELP)**, focused on developing essential skills such as Growth Mindset, self-management, Emotional Intelligence, Leadership effectiveness, Relationship building, Collaboration, Conflict resolution, and Feedback.
- Currently exploring the most recent advancements in AI, Large Language Models, and Prompt Engineering to explore their potential applications within our product and identify relevant use cases.
- Currently participating in a Kafka-related course to gain insights into its framework and proficiently utilize Kafka. As in future we may want to equip our product with Kafka to get streaming data.

Population Identification for Severe Asthma

(Ad hoc Request)

- Identified patients diagnosed with Asthma based on ICD-10 Codes
- Recognized patients who underwent Asthma Related Procedures via HCPCS codes
- Tracked patients receiving Asthma Related medication with NDC Codes
- Compiled data to create a Severe Asthma patient list for Medicare, Medicaid, and Commercial LOBs
- Processed approximately 3-4 billion raw data rows by combining multiple tables.

Intelligent PAC Provider Recommendation Engine

- Designed a holistic Recommender Framework to recommend not only the Best but the Right Provider to a patient, utilizing patient's health, Demographic, Socio-Economic profiles, and professional footprints of the provider based on location and patient's current health profile.
- **Techniques Used:** Created Patient Health Personas utilizing claims data, used unsupervised Deep Learning based clustering algorithm (**DEC**) to identify patient health personas. Performed t-Statistics neighborhood encoding (**t-SNE**) to visualize the identified clusters.
- **Rating and Ranking Framework:** Designed Multi-Variable rating framework to measure and rank-order based on consolidated performance of providers and Provider recommendation framework based on Patient's Need, location of patient, provider availability, patient's current health profile.
- **Impact Assessment:** Proposed and designed Impact Assessment framework to assess and estimate initial value of the Product.

Disease Prediction Model

Jun'17 – Sep'18

- Developed a Machine Learning model to identify the High-Risk members for a life-threatening disease.
- The developed solution was the first of its kind solution in the industry as it was able to identify the high-risk patients, with a 16x efficacy, even if patient is at home and model was utilizing only claim based data.
- **Techniques Used:** Utilized Feature Engineering derived clinically related features and Ensemble Model of finely tuned Random Forest (RF) and Gradient Boosting Machines (GBM) in the model.

Potential Abuser (of a Drug) Prediction

Oct'18 – Mar'19

- Developed a model to identify potential abusers for an addictive Drug using the features based on claims data and Pharmacy Data
- **Tools and Techniques Used:** Coded the complex interaction features in SAS EG and developed an Ensemble model of finely tuned RF and GBM to identify the patients at high risk of abuse of that Drug in R.
- Used Stratified Sampling technique to break the huge amount of data into 5 samples and used voting mechanism for final prediction out of predictions based on these samples

Covid-19 Detection Using Chest X-Ray Images

Runner-Up Solution: NEST Funninest DS Event

- Developed a **CNN** based Classification-Engine to classify X-ray Images into **Normal, Pneumonia** and **Covid-19** with class avg. recall 91% and avg. precision 94%.
- **Tools and Techniques Used:** Used **DenseNet121** CNN based architecture (429 layers, 121 CNN Blocks) with customized top layers for our problem. Took advantage of **Transfer Learning** by using pretrained weights on 14k images, freed initial 25 CNN blocks and trained the rest on our data. Used Class Prevalence adjusted class weights to give equal importance to each class to handle **Class Imbalance** problem. To **avoid overfitting** on training data used **Data Augmentation** techniques in-built in **TensorFlow**. For **Efficient Learning** and to avoid overfitting used TF's **LR scheduler** in **Adam optimizer** and **Keras** custom callback function to avoid training beyond 99.5% training accuracy. Used **TF's TensorBoard** to visualize model progress in run time and monitor and compare metrics in different iterations of model development.
- **Model Visualization:** Used recently proposed (2017) **Grad-Cam++** to visualize whether my model is focusing on the right areas of image or not while predicting a specific class. We visualized the heat map generated by Grad-Cam++, model was focusing on a single opacity in case of pneumonia and 3-4 opacities while predicting covid-19 and for normal it was focusing on parts other than lungs, and this observation was clinically verified.
- Proposed an Intelligent Covid-19 Progression Tracking framework using heat maps generated by **Grad-Cam++** in order to assess severity of Covid-19 and to track progression of Covid-19 for a patient.

DS Competency

Machine Learning	Neural Nets (ANN/CNN/RNN)	Sequence Models	Natural Language Extraction
Service Personalization	Bagging/Boosting	Model Interpretation	Medical Image Analytics

Techniques used

K-Means	Deep Embedded Clustering	Random Forest	XGboost	GBM	Ensemble Models
t-SNE	Information Value	Stacked-Auto-Encoder	Transfer Learning		Data Augmentation
Grad-Cam++	TF2.0's Object Detection API	Keras Custom Callbacks	TensorBoard		Keras

Language/ Tools Used

Python	CPython	Snowflake	DataBricks	Github Enterprise	Plotly Dash
R	R-Shiny	H2O	Spark	Zeppelin	SAS-EG
TensorFlow	Keras	Teradata			

Relevant Coursework

Deep Learning Specialization

Offered by Deeplearning.ai

Specialization designed to prepare learners to understand the capabilities, challenges, and consequences of Deep Learning. This is package of 5 courses Neural Networks and Deep Learning, Improving DNNs, Structuring Machine Learning projects, Introduction to CNNs, Sequence models.

TensorFlow in Practice Specialization

Offered by Deeplearning.ai and Google

Specialization designed to provide deeper understanding of Deep Learning, practice of understanding different type of layers, Callbacks, loss functions, optimizers using Frameworks provided by TensorFlow and Keras. Introduction to TF in AI, ML and Deep Learning, CNN in TF, NLP in TF, Sequence Models in TF were 4 courses.

AI for Medicine Specialization

Offered by Deeplearning.ai

Specialization design to provide practical experience of applying ML and Deep Learning to concrete problems in healthcare. AI for Medical Prognosis, AI for Medical Diagnosis and AI for Medical Treatment were 3 courses offered in this Specialization.

Machine Learning Engineering for Production (MLOps) Specialization

Offered by Deeplearning.ai

3 Course Specialization which focuses on Designing an ML production system end-to-end: project scoping, data needs, modeling strategies, and deployment requirements. This also focuses to let you learn to establish a model baseline, address concept drift, and prototype how to develop, deploy, and continuously improve a productionized ML application. In this I learnt to build data pipelines by gathering, cleaning, and validating datasets and applying best practices and progressive delivery techniques to maintain and monitor a continuously operating production system.

Practical Data Science Specialization

Offered by AWS and Deeplearning.ai

3 Course Specialization designed to understand and utilize services being offered by Amazon Web Services. In this course I learnt to Build, deploy, monitor, & operationalize end-to-end machine learning pipelines in AWS ecosystem, also learnt to label data using human-in-the-loop pipelines to improve model performance with human intelligence.

Relevant Certifications

- **AZ-900** – Microsoft Azure Fundamentals
- **AI-900** – Microsoft AI fundamentals
- **AI-102** – Designing and Implementing a Microsoft Azure AI solution.
- **DP-100** – Designing and Implementing a Data Science Solution on Azure
- **DP-090** – Implementing a Machine Learning Solution with Microsoft Azure Databricks
- **DP-203** – Data Engineering on Microsoft Azure

Appreciation & Awards

- Received well-recognized **“Master-Mind”** of Quarter for Contribution in Innovation.
- Received **“Best Team”** of the year for Enterprise-Level recognized, Life- threatening disease prediction model and its performance in pilot. This was well appreciated by client as well as senior leadership.
- Received **“BRAVO”** award for securing 1st rank in **“Predicting Adverse Events for Asthma Patients”** ML challenge organized by Optum.

Achievements/ Position of Responsibilities

- Provided exhaustive training on Python fundamentals to 350+ people from Optum as well as new campus hires in 2021
- **R Overview trainer** for newly recruited students from Different College Campuses in Optum.
- Received **B.H.U. Medal** for securing **1st position** at **B.Sc. (Hons.) Statistics Exam 2015**.
- Secured **AIR50** in **Joint Admission to M.Sc. (JAM)** of IITs 2015.
- Secured **AIR2** in PG Entrance test for M.Sc. in Statistics of BHU, 2015
- **INSPIRE Scholar**, under DST Govt. of India (Secured place in top 1% students in U.P. Board Intermediate (12th) Examination-2011)
- **Election Spectator** HEC Election 2017 Hall 11 I.I.T. Kanpur