Saurav Das

Data Scientist

LinkedIn Profile: https://in.linkedin.com/in/saurav-das-15bb0066
saurav_d234@@protonmail.com | +91-8013-289068 | +91-8013-289068
Open for remote work openings as an independent contractor
India based non US citizen | with no US work visa

My Portfolio

Work Experience

[Fresher] Looking to start a Data Science career, in progress.

I have completed numerous professional level projects, which are detailed below:

- **IBM Advanced Data Science Capstone Project (click for details).**
- **IBM AI Capstone Project with Deep Learning (click for details).**
- **University of Michigan: Applied Plotting, Charting & Data Representation in Python - Plotting Data Capstone Project (click for details).**
- **Google Data Analytics Capstone: Complete a Case Study (click for details).**

The projects mentioned above are readily shareable with recruiters and potential employers. In addition, I've successfully completed nine specializations and a plethora of courses, some of which involve projects with closed-source code or non-shareable licenses. For more comprehensive information, please refer to the [Education History] section.

I have also completed several guided [12 nos] projects of Coursera Project Networks [click for details].

Furthermore, in the interest of safeguarding the stakeholders' concerns, I am open to the idea of undertaking a test or on-boarding project. This would provide a practical opportunity for me to showcase my skills and capabilities, all with the aim of ensuring your convenience and confidence in my abilities.

[Other Experience] As Director / Stakeholder

A. K. Das LLC - November 2006 to Present (Over 17 Years)

I embarked on a remarkable journey in November 2006 by founding A. K. Das LLC, an enterprise that has since blossomed into a preeminent retail emporium specializing in bespoke clothing and premium dress materials. I have meticulously nurtured this endeavor into what it is today. At its core, my role as Director and Stakeholder has encompassed the following key achievements:

Inception and Growth: I initiated A. K. Das LLC from the ground up, architecting its foundations and fostering its exponential growth. I have transformed it into a distinguished establishment • **Leadership Excellence:** I have had the privilege of leading a dynamic team of nine individuals. This talented ensemble includes a proficient manager, a skilled accountant, and seven adept salespersons. Together, we have forged a harmonious synergy that fuels our success • **Online Expansion:** In 2020, I orchestrated the launch of our online store. This endeavor involved assembling and overseeing a team of accomplished temporary writers and creative designers. Their collective efforts were instrumental in effectively showcasing our diverse range of products and services to a wider audience. In sum, my tenure as Director and Stakeholder at A. K. Das LLC stands as a testament to my unwavering commitment to excellence, my adept leadership skills, and my keen sense of adaptability in an ever-evolving marketplace.

Soft Skills

[I possess a well-rounded set of Soft Skills, including:]

Communication: Proficiency in both written and verbal communication, allowing for clear and effective conveyance of ideas and information • **Strategic Decision-Making:** The ability to make well-informed and strategic decisions, considering various factors and potential outcomes • **Analytical Skills:** Strong analytical capabilities, enabling me to assess complex situations, identify patterns, and derive meaningful insights • **Adaptability:** A flexible and adaptive mindset that allows me to navigate change and new challenges with ease • **Creativity:** A creative approach to problem-solving, fostering innovative solutions to diverse problems • **Empathy:** A keen sense of empathy, which aids in understanding the perspectives and emotions of others, enhancing collaboration and interpersonal relationships • **Management Skills:** Proficient in management skills, including the effective oversight and delegation of tasks, ensuring smooth workflow and project success. These soft skills collectively contribute to my ability to excel in various professional and interpersonal situations, making me a valuable asset in both individual and team settings.

Relevant Technical Skills

The following list of skills is comprehensive, encompassing a wide range of competencies acquired from the various courses and specializations I have successfully completed. These skills, while not presented in a specific order, reflect my proficiency and versatility in addressing diverse challenges and tasks.

[I have a comprehensive skill set in Machine Learning, including:]

Python Programming: Proficiency in Python, a versatile language for machine learning • **Data Manipulation:** Expertise in Pandas and Numpy for efficient data manipulation • **Data Preprocessing:** Skilled in loading CVS and XLSX files, data-frame manipulation, data cleansing and exploratory data analysis, feature engineering and new feature generation • **Machine Learning Algorithms:** Able to distinguish between regression, classification, and clustering models clearly. Familiarity with a wide range of ML algorithms, including support vector machines, decision trees, random forests, gradient boosted trees, and various regression techniques like linear least-squares, ridge, lasso, polynomial and logistic regression · **Clustering:** Capable of implementing classification models like K-Nearest neighbors and hierarchical clustering • **Model Evaluation and Training:** Using Scikit-Learn to evaluate ML models effectively and cross-validation for model training • **Evaluation Metrics and Model Optimization:** Proficient in assessing model performance with confusion matrices, precision-recall and ROC curves, f1 score and other performance measures. Hyper-parameter tuning using GridSearchCV and RandomizedSearchCV functions. • **ML Engineering:** Able to create, save and evaluate (with new data) ML pipelines for robust data processing • **XGBoost:** Comprehensive knowledge on the XGBoost library a de facto industry standard library for machine learning • **Natural Language Processing:** Experienced in NLP using the Natural Language Toolkit (NLTK) for text manipulation, topic modeling, and document grouping • **Network Analysis:** Proficiency in NetworkX for network analysis, including centrality measurement and predicting network evolution • **R Programming:** Knowledge of R, R Markdown, and R-Studio, including knowledge of deplyr, readr, tidyr, ggplot2 etc packages. These skills collectively enable me to tackle a wide range of data analysis and machine learning tasks, making me a valuable asset for data-driven projects and problem-solving.

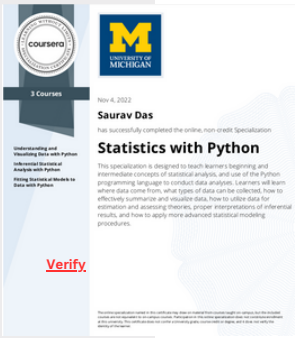
[I have a comprehensive skill set in Big Data Processing And ML Applications, including:]

IoT Data Processing: Proficient in handling and analyzing data generated by Internet of Things (IoT) devices • **Foundational Knowledge:** Developed foundational knowledge of Big Data technologies, including Apache Hadoop, MapReduce, Apache Spark, Spark SQL, and Kubernetes • **NoSQL Databases:** Competent in working with NoSQL databases, including inserting, updating, deleting, querying, indexing, aggregating, and sharding/partitioning data • **Database Experience:** Hands-on experience with NoSQL databases such as MongoDB, Apache Cassandra, and IBM Cloudant • **NoSQL Operations:** Capable of performing common tasks like CRUD (Create, Read, Update, Delete) operations in MongoDB and Cassandra • **ETL Processing:** Proficient in Extract, Transform, and Load (ETL) processing and machine learning model training and deployment using Apache Spark • **Apache Spark:** Experienced in using Apache Spark for massive parallel data processing, with a focus on improving analysis efficiency through partitioning and parallel analysis • **Spark Programming:** Applied Spark programming basics, including parallel programming for DataFrames, data sets, and Spark SQL • **Hadoop Ecosystem:** Understanding of Apache Hadoop architecture, ecosystem, and best practices, including components like HDFS, HBase, Spark, and MapReduce • **Cluster Management:** Demonstrated the ability to connect to Spark clusters, build ML pipelines, perform feature extraction and transformation, and manage model persistence • **Optimizing Spark:** Knowledgeable in optimizing Spark using utilities like Catalyst and Tungsten and proficient in using Spark's development and runtime environment options • **Machine Learning Pipelines:** Able to construct data analysis processes using Spark SQL and perform regression, classification, and clustering using SparkML. These skills encompass a comprehensive understanding of big data processing and machine learning applications for big data in various contexts, providing valuable expertise for data-driven decision-making and analysis.

Verify

[I have a comprehensive skill set in Data Visualization, including:]

Python Programming: Proficient in Python programming, a versatile language for data visualization and analysis • **Matplotlib & Seaborn:** Skilled in creating effective visualizations using Python libraries like Matplotlib and Seaborn to convey data insights • **Visualization Types:** Capable of producing various types of visualizations, including Box Plots, Run Charts, Line and Scatter Plots, Bar and Pie Charts, KDE and Histograms and Multidimensional Scaling plots • **Tableau Software:** Familiar with Tableau, a powerful data visualization tool, and proficient in using it to create interactive and insightful visualizations • **R Visualizations:** Knowledgeable about the options for generating visualizations in R, including libraries like ggplot2 and plotly, and able to leverage these tools for data visualization • **Tool Selection:** Able to identify the appropriate data visualization tool for specific tasks and understand their respective use cases • **Presentation Skills:** Competent in creating presentations that effectively communicate data-driven insights through visualization to enhance data storytelling and decision-making processes.



Verify

[I have a comprehensive skill set in Statistical Analysis, including:]

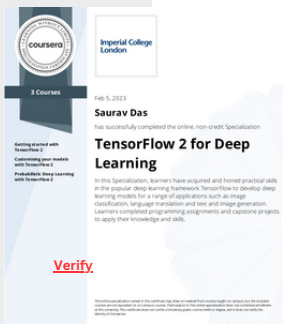
Sample Size Determination: Proficient in determining appropriate sample sizes for statistical analyses • **Sample Types:** Able to identify appropriate analytic techniques for both probability and non-probability samples • **Inferential Statistical Analysis:** Skilled in conducting inferential statistical analyses, numerical summaries and drawing meaningful conclusions from data • **Statistical Models:** Experienced in creating and applying statistical models to analyze data effectively • **Statistical Inference Methods:** Knowledgeable in various statistical inference techniques for hypothesis testing and parameter estimation • **Regression Models:** Capable of working with linear and logistic regression models, as well as multilevel regression models • **Bayesian Inference:** Familiar with Bayesian inference techniques for probabilistic modeling and analysis • **Statistical Tests:** Skilled in conducting statistical hypothesis testing, including the use of Student's t-test and Chi-squared test, and calculating confidence intervals • **Communication:** Proficient in communicating statistical ideas clearly and concisely to a diverse audience.

Verify

[I have a comprehensive skill set in Data Analysis, including:]

Data Formats: Proficient in working with JSON, XML, and databases (DBMS) for data storage and retrieval • **Web Data Acquisition:** Skilled in web scraping and using regular expressions to extract data from strings, with a deep understanding of web browser protocols • **API Integration:** Capable of retrieving data from websites and APIs using Python, including working with XML data and utilizing the Google Maps API for visualization • **Data Management:** Experienced in data cleansing, analysis, and organization using spreadsheets and metadata • **Data Collection:** Knowledgeable in ethical data collection practices and the importance of data ethics • **SQL Proficiency:** Competent in SQL for data query including joins from multiple tables, calculations, aggregation, and maintaining data integrity • **Data Quality:** Able to distinguish between biased and unbiased data and understand the functions and components of databases.

Verify



Verify

[I have a comprehensive skill set in Deep Learning, including:]

TensorFlow 2.0 and Keras: Proficiency in TensorFlow 2.0 and Keras for developing deep learning models • **PyTorch:** Experience with PyTorch for building and implementing deep learning algorithms • **Deep Neural Networks:** Building and optimizing various deep neural networks, including feed-forward networks, CNNs, RNN and LSTM • **Image Processing:** Utilization of Python, Pillow, and OpenCV for image processing and classification • **Computer Vision:** Expertise in computer vision and image recognition and classification • **Object Detection Models:** Familiarity with object detection models like Faster R-CNN, SSD, and YOLO • **Anomaly Detection and Time Series Forecasting:** Proficiency in anomaly detection and time series forecasting using deep learning • **Natural Language Processing (NLP):** Capability in applying deep learning to NLP tasks • **Unsupervised Learning:** Understanding and implementation of unsupervised deep learning models like auto-encoders and restricted Boltzmann machines • **Transfer Learning:** Proficiency in transfer learning techniques for reusing pre-trained models • **Model Optimization:** Tuning deep learning models using TensorFlow and PyTorch • **Deep Learning Pipelines:** Ability to create save and deploy end-to-end deep learning pipelines • **Real Data Applications:** Applying deep learning knowledge to improve models with real-world data • **TensorFlow Probability:** Familiarity with TensorFlow Probability for probabilistic modeling • **Generative Models:** Understanding and practical experience with generative models • **Probabilistic Programming Language (PRPL):** Competence in using probabilistic programming languages for modeling uncertainty • **End-to-End Workflow:** Ability to create end-to-end workflows for deep learning model development using TensorFlow and Pytorch • **Communication:** Demonstrated ability to effectively communicate and present outcomes of deep learning projects. These skills collectively make me proficient in various aspects of deep learning, enabling me to tackle complex projects and deliver meaningful results in artificial intelligence and computer vision domains.

[I possess a diverse range of General Computer Skills, including:]

Microsoft Windows: Expertise in using the Microsoft Windows operating system and its graphical user interface (GUI) shell • **Linux:** Operating knowledge of the Linux operating system (mainly Ubuntu and variants) and the bash shell • **Networking:** A general understanding of the TCP/IP protocol, subnets, routing, and related networking concepts • **Office Suites:** Proficiency in utilizing spreadsheet, word processing, and slideshow presentation software, using Microsoft Office and LibreOffice • **Version Control:** Proficient in Git and GitHub for efficient code versioning and collaboration • **Cloud Computing Environments:** Experience with cloud-based platforms like Google Colab, IBM Watson, and Paperspace, including IPython cloud notebook environments • **Cloud Storage:** Familiarity with cloud storage services such as Amazon S3 and IBM cloud storage • **Programming Languages:** Practical knowledge of several programming languages, including C/C++, C#, Python, R, and Scala • **Virtualization:** Technical expertise in virtualization technologies, including Vmware Workstation, ESXi, QEMU, VirtualBox, and Docker-based containers. These skills collectively represent a well-rounded and versatile skill set, enabling effective utilization of various software and platforms across different computing environments.

Education History

[Coursera - University of Michigan Specializations]

- **Python for Everybody Specialization. (click for details)**
- **Statistics with Python Specialization. (click for details)**
- **Applied Data Science with Python Specialization. (click for details)**

[Coursera - Imperial College London Specializations]

- **ICL Mathematics for Machine Learning Specialization. (click for details)**
- **ICL TensorFlow 2 for Deep Learning Specialization. (click for details)**

[Coursera & IBM Cognitive.ai - IBM Specializations & Professional Certificate]

- **IBM AI Engineering Professional Certificate. (click for details)**
- **Advanced Data Science with IBM Specialization. (click for details)**
- **NoSQL, Big Data, and Spark Foundations IBM Specialization. (click for details)**
- **IBM Cognitive.ai - Various Learning Paths And Courses (click for details)**

  Learning Paths: Applied machine learning With Python Level – II • Data Science Level -II • R Essentials • Big Data Foundations Level – I • Spark Level – II • Scala Programming for Data Science - Level II • Deep Learning Level – II

  Miscellaneous Courses: Statistics 101 • Mathematical Optimization for Business Problems

[Coursera - Google Professional Certificate]

- **Google Data Analytics Professional Certificate. (click for details)**

[Coursera - Miscellaneous Courses 5 nos.]

- **Miscellaneous Courses on Algorithms, Bio-Statistics, Scala and Git (click for details).**