ASHUTOSH SHUKLA

Address: A- 8 & 9, Sector 60, Noida, Uttar Pradesh 201301

E-Mail: ashukla.cms@gmail.com Contact No.: +91 9971471667

Experienced Senior Data Scientist /Architect with Team/People Management Skills in Various Domains

Profile Snapshot

- 13+ years of diverse academic and corporate experience in R&D as a scientist/researcher and 6+ year post Ph.D. experience across domains such as IOT, M2M, Automotive, Insurance/Financial, Retail, Software engineering, Quality, Sensors analytics, Big data Architect etc.
- Expertise in innovations/research and work closely with clients and delivery teams to find applications and business use cases for advanced analytical proficiencies. Filling patent in software engineering domain and published/presented/submitted more than 27 articles in various international/national journals/conferences and peer-reviewed conference proceedings.
- Experienced for solve challenging problems in the fields of Big data architect (specially distributed data bases and collaborative platform by utilizing Spark, Hadoop and Scala), Learning systems, Pattern recognition, Marketing Operation, Financial and other cyber-physical systems. Working knowledge in Big Data applications/solutions such as Mahout, Hive, HBase, Cassandra, Pig, no-SQL, etc. Knowledge of Hadoop/grid based programming for large scale problem solving.
- Expertise in active learning, Gaussian processes and Bayesian paradigm such as Bayesian network, Bayesian inference, Bayesian optimisation, Bayesian quadrature etc. Design and implementing algorithms is capable of understanding and acting upon big data that is complex and uncertain. I work to create modular numerical algorithms that speak the common language of probability theory.
- Experienced in the field of machine learning and robotics with emphasis on Bayesian learning, hierarchical/latent models, approximate inference, learning by demonstration, active learning and cognitive science (e.g. motor simulation view of social cognition).
- Experienced in implementation of Artificial Intelligence algorithms (Third generation Artificial Neural Networks (e.g.: Auto Encoders, Deep belief networks), Neuro-Fuzzy Networks, Advanced Kalman filtering (e.g.: Cubature Kalman Filtering, Information Filtering, Bayesian filtering, non-parametric filtering etc.), Dynamic Programming and Cognitive dynamic system)).
- Experienced in modelling issues such as concept drift examining, examining model compatibility, monitoring models and performing model comparison to get a most appropriate model among the compatible ones.
- Proven experience identifying, developing, and implementing technical solutions to enhance competitive market positioning; well-versed and hands-on expertise in applying Bayesian statistical and Markov Chain Monte Carlo (MCMC) models and methodologies.
- Experienced in building teams and partnering with global business stakeholders and technology teams to drive data analytics/science initiatives
- Experienced in Advanced quantitative methods relevant to modeling consumer experience in the digital world. Experience in web log mining for visitor segmentation, visitor behaviour modeling, common path analysis, conversion analysis, abandonment analysis, promotion analytics, buzz analysis, sentiment analysis, social networking analysis etc.
- Experience with analysis and optimization across marketing channels such as Digital Marketing (SEO/SEM, Display, Affiliates, Social Media), Direct Marketing (Email, Direct Mail), online (website) properties, CRM, and marketing automation tools.
- Extensive knowledge of R, Python, Scala, Matlab, FORTRAN 77, C, C++, Python, Maple, Mathematica, and others Productivity Applications: TEX LATEX, BibTEX, most common productivity packages (for Windows, and Linux platforms) Operating Systems: Microsoft Windows family, Linux and other UNIX variants.

Work Experience

HCL Technologies Ltd. [October 2014 to Present]
A-8 & 9, Sector 60, Noida, Uttar Pradesh 201301

Job Description: Senior Data Scientist / Senior Technical Architect (Engineering and R&D Services)

- Provide senior level guidance and advice on the development and implementation business analytics technology, including the use of predictive analytic models to achieve organizational and client goals. Oversee the development of experimental design software, enabling flexible design restrictions to position the company as one of the only firms capable of handling complex design restrictions.
- Created and currently managing an org of 10+ data scientist, data architect people that includes 2 agile development teams, the client QA team, and all Cloud Drive project management.
- Led the entire Cloud Drive team in establishing an agile development project and release process that successfully moved the team away from 1-2 big-bang-launches a year to 8 clients launching independently every 3-6 weeks.
- Created big-data innovative algorithm for Descriptive-statistics, Conversion, Hypothesis Testing, Missing value, Standardization and Machine learning algorithm in a distributed databases environment by utilizing Spark, R-Spark and Scala.
- Design and implementing algorithm for detecting anomaly in functioning of a switch/server/virtual machine/software subsystem based on collected statistical data. As well as use clustering algorithms to group switches/servers/virtual machines/storage systems based on various metrics to identify outliers.
- Design and implementing Software engineering solutions (product quality, developer rating, bug localization, bug privatizations, tester rating, support engineer rating, defect classification, resource allocation etc. by utilizing advance statistical and machine learning methods) for the deployment of various clients like CISCO, Mitel, ALU, Google, Boeing etc.
- Building algorithmic models for IOT, IVHM, HAR, M2M Analytics, Service Intelligence, Support Intelligence, and Product Intelligence.
- Conducting innovation workshops with clients on analytics and consulting on trends and emerging technologies in Analytics.
- Identifying and driving innovative projects in the area of sensor analytics and Internet of Things (IoT), using SCADA and operational data, to improve distribution center performance.
- Well-defined appointment process for data sciences talent and hired data scientists to build the data sciences competency within the organization.
- Leading a team to improve data quality and develop next generation business intelligence solutions through advanced visualization and automation.

Kvantum Marketing Insights Pvt. Ltd. [June 2013 to October 2014]
Tower-B, Cyber Park Gurgaon Haryana-122001

Job Description: Scientist (Applied Research)

- Design and implementing Retail solutions for various clients like Kimberly-Clark, TNF, Target, Ross Stores, Macys, etc.
- Led a team of analysts, technical consultants and project manager along with playing the role of a scientist to drive the implementations
- Implemented Media mix optimization and advertising response models, Marketing influence attribution in digital media and Measuring Synergistic Media Channel Performance in an Online Environment for various retail clients like Target, Kimberly-Clark, etc.
- Design and development of ongoing Target owned data feeds into data storage within Target's firewall Hypothesis models, influence algorithms, recursive density estimations and other baseline inputs needed. Attribution models for select action sets (FIATs, registry, RED-card events, and Flex Fulfilment options).
- Associated with global business partners and teams to drive "key" data sciences initiatives in the area of Social Listening (text mining), Pricing (forecasting) and Clickstream Analytics
- Worked with the global technology teams to adopt big data and open source tools within BI&A at Target India

Centre for Mathematical Sciences (CMS), Kerala [August 2011 to April 2013]
Department of Science and Technology (DST)
Job Description: Research Scientist (Assistant Professor),
- Research on Bayesian statistics, and machine learning algorithm.
- Developing life testing models for the Bayesian context especially in the situations where data available are insufficient, compound with (complex) censoring method
- Supervision of Ph.D. students, postdoctoral fellows and master students. Research project management.
- Conducted research on applying the method of maximum entropy to audio model for speech recognition.
- Invited by many Institute/university for presented paper in Bayesian statistics and life testing models.
- Prepared and delivered a graduate/mater-level course in different mathematical and statistical stream.

Stat-Math Unit, ISI-Delhi [December 2010 to March 2011]
Job Description: Visiting Faculty

- Conducted research on applying the Bayesian approach for estimating the parameters of hazard and reliability functions under the situations of asymmetric loss by utilizing MCMC simulation for load sharing models.
- Invited by many Institute/university for presented paper in Bayesian statistics and load sharing models.
- Prepared and delivered a graduate-level course in different mathematical and statistical stream.

Centre for Atmospheric Sciences, IIT-Delhi [July 2010 to October 2010]
Job Description: Postdoctoral Research Fellow
- My contribution, led to complete statistical analysis, developed stochastic models and novel decision regarding different circumstances of weather time series prediction in Bayesian context.
- Key Contributions:
- Three different classical statistical approaches, viz. singular value decomposition-based multiple regression, supervised principal component regression and canonical correlation analysis as well as Bayesian approach (utilizing exponential family of prior mainly for winter forecast) are applied on rainfall products obtained from eight global models.

Academic Credentials

PhD in Statistics [2009], B.H.U.-Varanasi, INDIA.
Specialization: Bayesian Statistics, Machine learning, and Computational Statistics.

Master of Science in Statistics [2002], B.H.U.-Varanasi, INDIA.
Specialization: Advance operation research and Demography
Title of Project: A sample survey on Quality of life, B.H.U. Employees

Bachelor of Science in Mathematics, Statistics, Physics [2000], V.B.S.P. University, INDIA.

Journal Publications, Talk, Invited Lectures, Workshop and Training Program

[1] Singh, B. P. and Shukla, Ashutosh (2009), Bayes Estimation of Shape Parameter of Classical Pareto distribution Using General Entropy loss Function, Journal of Scientific Research Banaras Hindu University Vol. 53, 147-162 ISSN: 0447-9483.

[2] Singh, B. P. and Shukla, Ashutosh (2009), Estimation of Shape Parameter of Classical Pareto Distribution by Shrinkage towards an Interval, Varahmihir, Journal of Mathematical Sciences Vol.2009, ISSN: 0972-7329.

[3] Shukla, Ashutosh (2009), Some Problems of Parameter Estimation of Classical Pareto distribution using Prior information, Unpublished Ph.D. thesis Banaras Hindu University Varanasi.

[4] Shukla, Ashutosh and Singh, B. P. (2012), Pareto Reliability Model: A Bayes Study under General Entropy Loss Function, Edited Volume "Application of Reliability Theory and Survival Analysis" 79-95, Bonfring Publication.

[5] Singh, B. P. and Shukla, Ashutosh, Bayes Estimation of Scale parameter of Classical Pareto distribution Under General Entropy loss Function, (submitted) Communication of Statistics Theory and Method.

[6] Shukla, Ashutosh, Bayesian estimation under general entropy loss for Pareto reliability model with application (possible submitted for publication), Journal of Bonfring Publication.

[7] Shukla, Ashutosh, Bayesian Estimation of Pareto Reliability Model under LINEX loss function (possible submitted for publication) IEEE TRANSACTIONS ON RELIABILITY.

[8] Shukla, Ashutosh, Bayesian Inference under Asymmetric loss from Pareto distribution with Application, (submitted for publication) Journal of Statistical planning and Inference-Elsevier.

[9] Shukla, Ashutosh, Reliability for the Load-sharing model: A Bayes Study under asymmetric loss (Manuscripts preparation for publication).

[10] Shukla, Ashutosh (2007), Estimation of Shape Parameter of Classical Pareto Distribution by Shrinkage towards an Interval, 29th Annual Conference of Indian Association for the Study of Population and National Seminar on Recent Statistical Techniques for data Analysis, Banaras Hindu University Varanasi.

[11] Shukla, Ashutosh (2008), Bayes Estimation of Shape Parameter of Classical Pareto distribution Using General Entropy loss Function, National Meet for research scholars in Mathematics and Statistics. December 6-10, 2008, Department of Mathematics and Statistics IIT Kanpur -208016.

[12] Shukla, Ashutosh (2008), Bayes Estimation of Scale parameter of Classical Pareto distribution Using General Entropy loss Function, VI International Symposium on Optimization and Statistics (ISOS-2008) December 29-31, 2008, Aligarh Muslim University, Aligarh, INDIA.

[13] Shukla, Ashutosh (2009), Bayes Estimation of Hazard function of Classical Pareto distribution Under General Entropy loss Function, Fourteenth Annual Conference of Gwalior Academy of Mathematical Sciences (GAMS) With Symposium on Computational Mathematics and its Application to Engineering, Management and Biology July 17-19, 2009.

[14] Shukla, Ashutosh (2009), Bayes Estimation of Reliability function of Classical Pareto distribution Under General Entropy loss Function, Eighteenth International Conference of Forum for Interdisciplinary Mathematics on Interdisciplinary mathematical and statistical Techniques, IMST 2009-FIM 18 August 2-4, 2009.

[15] Shukla, Ashutosh (2009), Bayes Estimation of Pareto life time model under LINEX loss Function, Seventh International Triennial Calcutta Symposium on Probability and Statistics December 28-31, 2009 INDIA.

[16] Shukla, Ashutosh (2010), Bayes Estimation of Scale Parameter of Classical Pareto Distribution under LINEX loss function using MCMC Technique, Article published in the Souvenir of International Conference on Statistics, Probability, Operations Research, Computer Science and Allied Areas in Conjunction with 8th IISA joint Statistical meeting and 29th annual Convention of ISPS held at Visakhapatnam, INDIA from 4th to 8th, January 2010.

[17] Shukla, Ashutosh (2010), Discriminating Among Different Loss Function in Pareto Income Model: A Bayes Study, International Conference on Optimization and its applications, Banaras Hindu University Varanasi February 16-18, 2010, (ICOIA-BHU-2010).

[18] Shukla, Ashutosh (2010), Bayesian Estimation of life time parameters of Generalized Pareto Model under Asymmetric loss function, Article published in the Souvenir of International conference on development and applications of statistics in emerging areas of science technology, University of Jammu INDIA. December 8-10, 2010.

[19] Shukla, Ashutosh (2011), Bayesian Estimation of Pareto Reliability Model under LINEX loss function, International Conference on "New Trends in Life Testing, Bayesian Inference, Sampling Theory, Biostatistics, Bioinformatics and Computer Applications, Banaras Hindu University Varanasi-221005.

[20] Shukla, Ashutosh (2011), Bayesian estimation of Pareto life time Model under Asymmetric loss function, Invited Talk; Stat-Math Unit, Indian Statistical Institute (ISI) Delhi, New Delhi. February 23, 2011, INDIA.

[21] Shukla, Ashutosh (2012), Bayesian Estimation for the Reliability of Load-Sharing systems under Asymmetric loss, National Conference On Recent Developments In The Applications Of Reliability Theory And Survival Analysis, February 02-03, 2012,Pondicherry University, Puducherry-605014.

[22] Shukla, Ashutosh (2012), Reliability of k-components Load-Sharing systems: A Bayes study under asymmetric loss, New Dimension in Statistics: Concepts and Issues, Banaras Hindu University Varanasi-221005, March 02-04, 2012, NDSCI-2012.

[23] Shukla, Ashutosh (2013), Bayesian Inference under Asymmetric loss from Pareto distribution with Application, ISBA Regional Meeting and International Workshop/Conference on Bayesian Theory and Applications (IWCBTA), Banaras Hindu University Varanasi-221005, January 6-10, 2013.

[24] Shukla, Ashutosh et.al (2015), Deep Learning Model for Near Perfect Defect Removal System over Multiple Release Cycles, submitted the following paper to MSR 2015.

[25] Shukla, Ashutosh et.al (2015), Bayesian Approach to Grade Software Developers based on Prior Measurable Parameters, submitted to the ISEC 2016.

[26] Shukla, Ashutosh (2015), Decision theoretic Bayesian Approach of testing model for improving test efficiency submitted to the ISEC 2016.

[27] Symposium and Workshop on Statistical Methods in Computational Biology, Mahila Mahavidyalaya Banaras Hindu University 28-29 December 2007.

[28] Advanced Training Program (ATP) on Bayesian Statistics and Computation Department of Statistics and DST Centre for Interdisciplinary Mathematical sciences, Banaras Hindu University Varanasi -221005 18 February-01 March 2008.

[29] 2008 SERC School on Matrix Variable Calculus and Statistical Distribution theory and applications in Data Analysis, Model Building and Astrophysics Problems to be held at the Centre for Mathematical Sciences Pala Campus Kerala, India, from 15th April 2008 to 17th May 2008.

[30] Workshop cum Training Program on Stochastic Models and Their Applications during September 22 to October 04, 2008, Department of Statistics and DST Centre for Interdisciplinary Mathematical sciences, Banaras Hindu University.

Project

1. Development and application on extended range forecast system for climate risk management in agriculture. [July 10 to Nov10] Development of a Climate forecast system (monthly to seasonal) at met-subdivision level (higher spatial resolution) along with six homogeneous regions and India as a whole. Experimental real-time extended range prediction of rainfall and temperature in monthly scale with seasonal outlook. End to end application of these climate forecast products in agriculture through 9AUs and feedback from end users (prospective farmers). The variability of Indian summer monsoon rainfall (ISMR; both at inter-annual as well as intra-seasonal timescales) has a direct impact on various sectors of public interest and economy such as agriculture, water resource, etc. So, the need of real-time extended range forecast system of the ISMR (monthly to seasonal scale) is not overstated. The present study is aimed at developing such a forecast system to predict rainfall (monthly and seasonal mean) one month in advance over 34 meteorological subdivisions of India for climate risk management in agriculture.

2. Media mix optimization and advertising response models. [June 2013 to September 2013]
    The intent of this project is to utilize the prearranged customer information, discover the new knowledge to improve decision making in marketing. The consumer is investing significant amounts of capital in diverse media channels, both digital and traditional (like TV, direct mail or Press) to improve his imaged or sales.

3. Campaign Attribution and Campaign Influence in digital media:

    This project is aimed to understand the offline impact (revenue/brand value/butts in seats/phone calls etc.) driven by online marketing and advertising. Working with cross functional teams in a fast paced multidisciplinary environment, specializing in big data, machine learning, artificial intelligence, statistics and Bayesian network to develop meaningful and valuable insights into data that can come from a variety of channels such as sales, marketing and clickstream.

4. Marketing influence attribution in digital media. [October 2014 to September 2013]

    The intent of this project is to illustrate the possibility to deliver a mathematical model that attributes the impact a sample guest and Client actions independently and in concrete have on changing guest value profit. This project provides tools to attribute guest influence to combined store and target.com sales and margin. To explore this I oversee Client core sales data on daily basis (online and offline) for serving algorithms and the optimization platform. Researched and created Machine learning methods for identifying features and dimension reduction by using advance data science technique.

5. Measuring Synergistic Media Channel Performance in an Online Environment. [April 2014 to October 2014]
    The intent of this project is detecting the direct effects of digital media (i.e. Facebook, google etc.) on sales and market share, as well as their potential synergy effects with large media (i.e. TV, magazines, newspapers and radio). In this project, I also quantify the extent of a potential self-selection in the advertising elasticity's of the different media. The results of this analysis will help to re-evaluate budget allocation in a context of smaller and less researched media.

6. Product Sustenance and Support (PSS) [October 2014 to June 2015].

    This project is aimed at improving and customizing authentication strategy by managing and supporting product lines for *Fortune* 500 companies in the U.S. and Europe, providing L1 and L2 product support, sustenance engineering, and migration services, and revolutionizing tech support and product sustenance using product intelligence.

    -   Intelligent Tech Support (ITS): Predictive analytics driven measurement of Value-Per-Interaction, Sophisticated big data mining technologies employed to generate insight, APIs for easy integration, Helps customer service representatives to reduce calls and escalations and improve CSAT by up to 30%. We also used analytics to identify solutions from chats with customers. Therefore, we were able to reduce call volumes and Mean Time To Resolution (MTTR) by 30%. Building an analytics based correlation engine which found automated solutions. This resulted in the improvement of mean response time by 15% and MTTR by 8%.
    -   Intelligent Sustenance Engineering (ISE): Helps to identify critical areas that the team can concentrate on during transition. Recommends testing strategies to facilitate quicker independence for the testing teams. Provides automated dashboards that help program managers identify bottlenecks in sustenance engineering. Escalation management tools streamline the escalation of tickets by verifying the process and allocating tickets to the right teams automatically. Helps prioritize bugs based on tech support and cost to company. Automatically sets up environment/configuration in which the bug was discovered, for faster debugging. Testing recommends the tests that need to be performed based on previous testing patterns and code changes. Recommends the tests that need to be performed for optimum coverage of parameters like environments, criticality, etc. Release Management tools efficiently co-ordinate with the teams to ensure smooth release cycles.

7. Cognitive Engineering & Support System (July 2015 to Present).

    This project is aimed at improving and growth of Cognitive Engineering & Support System systems that support of human work. Experience tells us that effective sense-making and decision support systems cannot be designed by an engineer's intuition alone. The design of information technologies should be informed by knowledge of how people think and act in the context of their work environment. Clearly, information technology design teams must be knowledgeable about computer software and hardware, but if the technologies fail to support cognitive functions, they will be rejected in the workplace and marketplace.

    -   Product Intelligence (PI): Designed solution for PI that help companies address the business challenges arising because of the increasing scope and design complexity of their products. Focused analytics solution for the product lifecycle. Setting up engineering telemetry data collection, processing, analysis and presentation of Greenfield and existing systems. Designed solution for PI that has the ability to learn past execution history (especially the test execution strategy) and perform sustenance activities such as hotfix releases and service pack releases efficiently with limited resources and knowledge base. Designed Failure prediction that helps companies plan predictive maintenance more effectively, leading to reduction in field support costs and product downtime. Bug and enhancement request prioritization, Optimize regression/other test suites by knowing various features/modules used by customers. Software defect localization and duplicate defect reduction.

    -   Big Data Analytics in M2M: Machine-to-Machine (M2M) means that no human intervention is required whilst devices are communicating with each other. It refers to technologies that allow both wireless and wired systems to communicate with other like devices. M2M is considered an integral part of Internet of Things (IoT). Designed solution for Vibration monitoring, this is the process of monitoring a parameter of condition in machinery (vibration, temperature etc.), in order to identify a significant change which is indicative of a developing fault. It is a major component of predictive maintenance), Thermal Imaging solution, ultrasonic monitoring solution and Human activity recognitions (HAR) solutions.

    -   PANGEA: Pangea is a distributed and collaborative analytics platform built on industry leading big data technologies that brings the power of machine learning to simplify the complexities in deriving insights from disparate at sources. Created distributed database innovative algorithm for Descriptive-statistics (24 functions), Conversion (categorical to numerical and vice-versa), Hypothesis Testing (10+ functions), Missing value, Standardization (Z-score and min-max normalization) and Machine learning algorithm (decision tree, random forest, naïve bayes, survival analysis etc.) in a distributed databases environment by utilizing Spark, R-Spark and Scala.

- Intelligent machine device analytics (IMDA): Induction motor-drive systems are used throughout industry. In many applications they are the motor-drives of choice. They are used in a wide variety of motor-drive applications such as industrial plant control, propulsion systems, and medical diagnostic equipment. Failure of such motor-drive systems have serious impact on the equipment they are part of including shutdown of the larger system. In industrial applications the impact can be dangerous and costly, but in medical applications such failure can even have fatal effects. Accurate diagnostics and fault predictive will increase the reliability of motor-drive systems and minimize the problem of their failure in the field. A dual track approach is proposed to developing a comprehensive diagnostic system for motor-drive systems. I address two fundamental problems in developing such a diagnostic system, namely the difficulty and cost of obtaining large amounts of mtoor-drive system failure data and the difficulty in developing accurate, robust, fast, and effective diagnostic techniques. The dual track is a combination of time stepping coupled finite element state space (TSCFE-SS) techniques for generating accurate, but inexpensive, fault data with a reconstructed phase space based approach for modeling system trajectories for creating effective fault signatures.