

# Dibyendu Chakrabarti

Double PhD, Computer Science – San Francisco Bay Area  
📞 (201) 259 9759 • ✉️ dibyendu.chakrabarti@gmail.com  
🌐 <https://www.linkedin.com/in/dibyenduc>

## A technical leader with a different attitude

---

- A Playing Coach, definitely not a Non-Playing Captain
- Currently working on several Deep Learning (DL) techniques in Computer Vision/NLP
- Hands on individual contributor in distributed computing: TensorFlow, PyTorch, MXNet, Scala, Python/Apache Spark, Scalding, Hive, Impala and Java
- Deep understanding of traditional Machine Learning Algorithms: Anomaly Detection, Supervised Learning (Classification/Regression), Unsupervised Learning (Clustering), Collaborative Filtering (CF), Dimensionality Reduction, Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Topic modeling (LDA)

## Recognition from U. S. Citizenship and Immigration Services [USCIS]

---

- US Citizen
- Earlier:
  - Green Card in the EB-1 Extraordinary Ability Category [Extraordinary ability in the sciences]
  - O-1A visa for individuals with an extraordinary ability in the sciences

## At a glance

---

- Double PhD and several years of work and research experience
  - Applied Machine Learning, New Jersey Institute of Technology, Newark, NJ, USA
  - Cryptography & Network Security, Indian Statistical Institute, Calcutta, India
- Experienced programmer with deep knowledge of algorithms
- Experienced researcher in computer science with several publications and more than 200 Google Scholar citations till date
- Editor and reviewer of prestigious journals/conferences. Made presentations at international conferences and visited foreign universities as a scholar

## Education

---

<b>New Jersey Institute of Technology</b> <i>PhD, Computer Science, Simulation Methodology and Stochastic Modeling</i>	<b>Newark</b>
<b>Indian Statistical Institute</b> <i>PhD, Computer Science, Cryptography/Combinatorics/Sensor Network</i>	<b>Calcutta</b>
<b>Indian Statistical Institute</b> <i>MS, Computer Science, Specialized in Graph Algorithms</i>	<b>Calcutta</b>

## Employment history

---

<b>Self-employed</b> <i>Personal Project for a Stealth Mode Startup</i>	<b>South Bay Area, USA</b> 2020–Till date
<b>Williams-Sonoma</b> <i>Principal ML Engineer (Director level)</i>	<b>South Bay Area, USA</b> 2019–2020
<b>Experian</b> <i>Senior Data Scientist</i>	<b>South Bay Area, USA</b> 2016–2019
<b>Verizon</b> <i>Principal Member of Technical Staff</i>	<b>South Bay Area, USA</b> 2014–2016
<b>Wolters Kluwer</b> <i>Computer Scientist (Manager, Customer Insights)</i>	<b>New York, USA</b> 2012–2014

## Skillset

---

- **MACHINE LEARNING AND LARGE SCALE DATA ANALYSIS:** Supervised Learning (Classification Techniques and Regression: Logistic Regression, Generalized Linear Model (GLM), Naïve Bayes, Support Vector Machines (SVM), Random Forest), Unsupervised Learning (Clustering Techniques: K-Means, Non-parametric, Spectral, Graph-based), Collaborative Filtering (CF), Dimensionality Reduction, Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Distributed systems, Distributed computing: Apache Spark, Scalding, Hadoop, MapReduce, NoSQL, Hive, Impala.
- **PROGRAMMING:** Independent analysis, design and implementation of algorithms, data structures and programs, software development, testing and debugging. Knowledge of C/C++, Java, Scala, Matlab, Mathematica, Python, R, SQL, Shell scripting in UNIX, several RDBMS' (Oracle, SQL Server) and OS' (OS X, Unix/Linux, Windows).
- **QUANTITATIVE:** Mathematical, statistical and stochastic modeling, analysis of large datasets, optimization techniques, time series models, numerical methods, probability, statistics, partial differential equation, stochastic differential equation, finite difference methods, regression and statistical analysis, stochastic process and Brownian motion, Monte Carlo simulation and variance reduction techniques.
- **Cryptography:** Cryptographic protocols: Homomorphic Encryption, group signatures, blind signatures, anonymous credential management, commitment schemes, zero-knowledge proofs, proof of knowledge, ZK-SNARK, oblivious transfer, secure multiparty computation, Oblivious RAM, private set intersections, private information retrieval. Blockchains: Hash functions, signature schemes, zero-knowledge proofs, consensus algorithms. Bitcoin: Transactions, blocks, mining, scripting, attacks on mining. Proof of work, proof of stake, proof of burn, proof of storage. Smart contracts Privacy issues: Anonymity, mixing techniques, privacy with ZK-Snarks. Permissioned blockchains: Distributed consensus, sharing algorithms, privacy issues. Scaling issues: Lightning networks, Payment networks. Platforms and ledgers: Ethereum, Ripple, Hyperledger, Algorand

## Recent project highlights

---

### Anonymous auction on Blockchain.....

The objective of this work is to hold an auction maintaining the anonymity of the bidders and the confidentiality of the bids. We take advantage of the auditability feature inherent to blockchain technologies and rely on the assumption of the existence of an underlying privacy preserving

blockchain platform. The design of such a protocol should follow a generic approach and use well-known cryptographic primitives such as commitment scheme, public key encryption and ring signature. The anonymity of the bidders may be achieved using a Designated Identity Verifier Ring Signature technique.

Blockchain platform: Ethereum, Auction Smart Contract: Solidity, Simulation of the main auction protocol: Python

### **Fine-Grained Visual-Categorization (FGVC)**.....

As online shopping and retail AI become ubiquitous in our daily life, it is imperative for computer vision systems to automatically and accurately recognize products based on images at the stock keeping unit (SKU) level. However, this still remains a challenging problem since there is a large number of SKU-level categories, many of which are fine-grained, with very subtle differences that cannot be easily distinguished. At the same time, images of the same product or SKU can often look different under different conditions (e.g., user generated content vs professional generated content). I am trying to develop novel algorithms to tackle this interesting problem.

### **Recommender Systems**.....

- o Deep Learning Recommendation Model (DLRM) for Personalization and Recommendation Systems
- o Matrix Factorization and Rating Prediction with Autoencoders
- o Personalized Ranking and Neural Collaborative Filtering
- o Sequence-Aware/Feature-Rich Recommender Systems
- o Factorization Machines/Deep Factorization Machines

### **AI techniques in Medicine/AI for Medicine Specialization**.....

- o AI for Medical Diagnosis
- o AI for Medical Prognosis
- o AI For Medical Treatment

### **Deep Learning Techniques/Deep Learning Specialization**.....

- o Neural Networks and Deep Learning
- o Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
- o Structuring Machine Learning Projects
- o Convolutional Neural Networks
- o Sequence Models

## **Past projects**

---

### **Experian**.....

- o Graph analytics for identity resolution [Use case: Entity resolution]
- o PMML: Development and management of several classification models in the portable PMML format [Use case: Demonstration of modeling capability in a product called PowerCurve]
- o Cybersecurity and IoT: Design and implementation of several attack-specific Anomaly Detection algorithms; Streaming model for an attack-agnostic Anomaly Detection algorithm [Use case: Design and detection of various attack vectors for a class of IoT devices]
- o Design of efficient (sub-linear) nearest neighbor finding algorithm [Use case: Entity resolution]
- o Design of high performance string clustering algorithms [Use case: Entity resolution]
- o Very large scale univariate/multivariate anomaly detection in user behavior [Use case: Fraud detection]
- o Definition and implementation of a new estimator [Use case: Performance metric]

- Implementation of a similarity digest scheme viz., Ssdeep [Use case: Comparison of event descriptions/Entity resolution]

## Verizon.....

- Development of algorithms to discover anomalies, trends, and patterns related to network performance in near real-time
- Analysis of DDOS attacks: Detection/Prediction of attacks using Statistical Techniques
- Scalable algorithms for clustering, univariate and multivariate anomaly detection using Scala and Apache Spark.
  - Simple Gaussian model
  - Multivariate Gaussian model
- SMS Spam Detection using Machine Learning Approach (Naïve Bayes)
- Latent Dirichlet Allocation (LDA) for Text Clustering
  - Built topic models using LDA from large number of text documents
  - Dimension reduction using the topics as features
  - Clustering the documents using the above feature space

## Publications

---

### Journal Articles

1. A key pre-distribution scheme for wireless sensor networks: merging blocks in combinatorial design, International Journal of Information Security, vol. 5, Issue 2, pp. 105–114, April, 2006.
2. Clique size in sensor networks with key pre-distribution based on transversal design, International Journal of Distributed Sensor Networks, vol. 1, Issues 3–4, pp. 345–354, 2005.

### Refereed Conferences

1. Application of Simulation in Designing a Secure Grid of Distributed Wireless Sensor Network, Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2009), Hong Kong.
2. Combinatorial structures for design of wireless sensor networks, Proceedings of 4th International Conference on Applied Cryptography and Network Security, ACNS 2006, LNCS 3989, pp. 365–374, 2006.
3. A key pre-distribution scheme for wireless sensor networks: merging blocks in combinatorial design, Proceedings of 8th Information Security Conference, ISC 2005, LNCS 3650, pp. 89–103, 2005.
4. Quantum Weight Decision Algorithm of a Boolean Function, Proceedings of ERATO conference on Quantum Information Science 2005, pp. 127–128, EQIS 2005, Japan.
5. A hybrid design of key pre-distribution scheme for wireless sensor networks, Proceedings of 1st International Conference on Information Systems Security, ICISS 2005, LNCS 3803, pp. 228–238, 2005.
6. Clique size in sensor networks with key pre-distribution based on transversal design, Proceedings of 7th International Workshop on Distributed Computing, IWDC 2005, LNCS 3741, pp. 329–337, 2005.

### Dissertations

1. Applications of stochastic simulation in two-stage multiple comparisons with the best problem and time average variance constant estimation, New Jersey Institute of Technology, 2011.
2. Applications of combinatorial designs in key pre-distribution in sensor networks, Indian Statistical Institute, Kolkata, 2007.