

PERSONAL PROFILE

Name: Anmol Batra
Address: #####, India
E-mail: #####@yahoo.com
Gender: Male
Date Of Birth: 31st May, 1987
Contact No:
Notice Period: 5 Days

OVERVIEW

I have been working **as a Data Scientist for 5.5 years** now. I have lead projects related to predictive modeling, recommender systems and anomaly detection using Supervised ML.

In all, I have worked as a **Data Analyst for 10 years** with diverse exposure. I have been largely responsible for data engineering tasks that included Data Mining, Data Cleansing, Descriptive Analytics, MIS Automation, Audit Reconciliations, CAAT, Insights Generation and Data Pipeline Automation.

I have been a **Project lead/Manger from past 5 years now**. And I have lead projects in both Data Sciences and Data Engineering. I have spent **7 years in Telecommunications industry and 2 years in Hospitality and Tourism**.

SKILL SET

- ✓ Programming Languages: Python, R, SQL, VBA
- ✓ Database/Analytics Tools: Amazon SageMaker, Dataiku, Jupyter, MS Access, Tableau
- ✓ Big Data: Hadoop, Apache Spark
- ✓ Project Management: Jira
- ✓ Data Analysis Skill set: Interactive Dashboards, Benchmarking, Data Ingestion, Revenue Assurance, Consumer Insights, Data Visualization, MIS Automation, ETL, Campaign Management Analytics
- ✓ Data Sciences Skill set: Recommender systems, Anomaly detection, Classification Models, Natural Language Processing, Mathematical Optimization

COURSES AND CERTIFICATIONS



[Microsoft Technology Associate](#)
 Database Fundamentals: T-SQL
 Level: Associate | Grade: 86%



[Machine Learning](#)
 offered by Coursera
 Level: Intermediate | Grade: 93%



[Applied Machine Learning in Python](#)
 offered by Coursera
 Level: Intermediate | Grade: 91%



Six Sigma - White Belt
 offered by: Ernst and Young
 Level: Associate

PROJECTS

Data Sciences:

Probability of Hotel Booking Cancellation: The Australian Hotel Services Reseller needed a methodology to reduce loss due to booking cancellation for its Partner Hotels. As a solution, a classification model is being built which can score each incoming Hotel booking for its probability of Cancellation later.

Project Type: Classification Model

Client: Hinstant.com

Duration: September 2019 – Present

Tools Used: Amazon SageMaker, MySQL, UNIX Crontab

Libraries Used: pandas, matplotlib, scikit-learn, numpy

Techniques Used: Logistic Regression, Regularization, Feature Scaling, AUC Optimization, GridSearch

Role: Project Manager/Data Scientist

Responsibilities:

- ✓ Consulting
- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Task Allocation
- ✓ Feature Engineering
- ✓ Exploratory Data Analysis
- ✓ Python Scripting for Algorithm Selection, Feature Scaling and Model Optimization
- ✓ Unix Cronjob for Data Pipeline Automation
- ✓ SQL Scripting for Feature Creation/Selection/Identification

Recommender System for Tourism Industry: The Australian Hotel Services Reseller needed to improve his Deal Conversion Rate against the impressions generated on email and on mobile application. As a solution a Hybrid recommender system (combination of Supervised & Unsupervised Machine learning) has been developed which would generate recommendations bases the historic user deals and independent clusters of users and products. A genetic algorithm would run further on top of Recommender system to optimize selection of scored Products.

Project Type: Recommender System

Client: Hinstant.com

Duration: March 2019 to September 2019

Tools Used: Amazon SageMaker, MySQL, UNIX Crontab

Libraries Used: Pandas, Seaborn, NumPy, LightFM

Techniques Used: Genetic Algorithm, LightFM, Feature Scaling, Recursive Feature Elimination, Bootstrapping, Precision at K.

Role: Project Manager/Data Scientist

Responsibilities:

- ✓ Consulting
- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Task Allocation
- ✓ Feature Engineering
- ✓ Python Scripting for Algorithm Selection, Feature Scaling and Model Optimization
- ✓ Unix Cronjob for Data Pipeline Automation
- ✓ SQL Scripting for Feature Creation/Selection/Identification

Predictive Analytics for Sporting Events: The Australian sports content aggregator was looking for a multiclass probabilistic model which could score probability of Team A/Team B/Draw before & during the match. The model uses data related to Betting odds/Social Media/Match Historics to generate event based real-time probability numbers.

Project Type: Classification Model

Client: Sportsflash.com.au

Duration: June 2018 to Sep 2018

Tools Used: SQL, Python, Selenium, UNIX

Libraries Used: pandas, matplotlib, scikitlearn, numpy

Techniques Used: Gradient Booster, Recursive Feature Elimination, Random Forest Feature Ranking, Sentiment Analysis, Grid Search, Hyper Parameter tuning, MapReduce

Role: Project Manager/Data Scientist

Responsibilities:

- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Task Allocation
- ✓ Feature Engineering
- ✓ Python Scripting for Algorithm Selection and Model Optimization

- ✓ SQL Scripting for Feature Creation/Selection/Identification

Sentiment Analysis for Air Purification Market: The client in consumer electronics domain was looking for insights into a fast growing product in Indian Consumer Market – Air Purifiers. The insights project was driven by Sentiment Analysis of social media data(i.e. Tweets/Comments/Status Updates/Product reviews) using R. The project also included a paid review classifier which would flag reviews as paid or genuine

Project Type: Classification Model, Consumer Insights

Duration: Jan 2018 to May 2018

Tools Used: Python, Microsoft Power Point

Techniques used: SVM Linear, Document Term Matrix, Lexicon Token Analysis, N-gram distribution, Sentiment Polarity scoring

Role: Project Manager/Data Analyst

Responsibilities:

- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Task Allocation
- ✓ Classification modeling
- ✓ Feature Engineering
- ✓ Python Scripting for Natural Language Processing, Classification
- ✓ Power Point Presentation

User SIM card fraud: Mobile Recharge Retailers would indulge in fraudulent sale of Aircel SIM cards. i.e. They would misuse received user documents to issue SIM cards to others & influence Users to buy SIM cards when approached for recharge owing to high Margin. To cater to the above 2 problems two interconnected models were built using Multivariate Gaussian Model which would score all retailers across India for their probability of involvement in above frauds.

Project Type: Anomaly detection

Duration: February 2016 to November 2016

Tools Used: R, Excel, Power Point

Technique used: Correlation Matrix, Anomaly detection, Multivariate Gaussian, Artificial Data Synthesis

Role: Project Lead/Data Scientist

Responsibilities:

- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Feature Engineering
- ✓ Task Allocation
- ✓ ETL Process creation

Reporting & Analytics:

IPEDS University Ranking System: My Client Mckinsey was looking for a Tableau based interactive Dashboard which could dynamically Rank Universities across United States bases weighted means. The Client wanted weights choice decision in the hands of the Mckinsey Consultant

Project Type: Tableau Dashboard

Duration: October 2019 to Jan 2019

Tools Used: MS Access(VBA), Access Forms, Tableau 2019.4, SQL

Technique used: Weighted Average, Rank Function, Geographic charts, Heatmap

Role: Project Manager/Business Analyst

Responsibilities:

- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Solution Architecture
- ✓ Task Allocation
- ✓ Wireframe Design

Data Ingestion for Retail Analytics Product: My client BCG needed multiple automation pipelines from its end-clients(i.e.Global retail giants) connecting to its Proprietary tool, Retail Catalyst. The client also needed automation of its existing monthly slides deck which was published to its end clients.

Project Type: Proprietary Tool

Duration: Jan 2018 to April 2018

Tools Used: Python, Dynamic SQL, DataIKU, Powerpoint, Apache Spark, Hadoop, IBM Thinkcell

Role: Individual Contributor

Responsibilities:

- ✓ Apache SQL Scripting for ETL
- ✓ Data Cleansing
- ✓ Data Lake pipelines
- ✓ Data Ingestion
- ✓ Python functions for Visual Recipes in DataIKU
- ✓ Thinkcell chart design
- ✓ DataIKU Visual Recipes
- ✓ Deployment and Scheduling

Reconciliation tool for Market Mix Modeling: My client BCG needed an Automation of existing Excel bases Reconciliation methodology for Marketing Budget Data. The Excel(VBA) based automation used to identify the gaps between the data uploaded in the Cloud Based tool against base data received from the 3rd Party service providers.

Project Type: Report Automation

Duration: March 2017 to June 2017

Tools Used: MS Excel(VBA), KNIME

Technique used: Recursion, Formula Arrays

Role: Project Lead

Responsibilities:

- ✓ Requirement Gathering
- ✓ Project Planning and Execution
- ✓ Task Allocation
- ✓ ETL Process creation
- ✓ VBA Scripting

Sequential Billing Activity: My Company Centurylink needed an interactive dashboard which could be used to track count/revenue new Business Customers on-boarding, exits, decrease/increase in revenue/usage on Month on Month bases.

Project Type: Tableau Dashboard

Duration: March 2013 to Oct 2013

Tools Used: MS Excel, Tableau, Oracle SQL

Technique used: Weighted Average, Rank Function

Role: Individual Contributor

Responsibilities:

- ✓ Requirement Gathering
- ✓ Solution Architecture
- ✓ Wireframe Design
- ✓ Deployment and Scheduling
- ✓ Data Mining
- ✓ Tableau Dashboard design

End-to-End Automation of a BI Tool: My client McKinsey, had developed a Silverlight based Analytical solution for pricing. As an Analyst I was responsible for configuring the Solution package as per client needs and setting up an End-to-End Automation for Consultant to analyze the data. I was also responsible for understanding client database to build water-fall/Scatter-plot models and programming MS Access to automate it.

Project Type: Data Ingestion & Data Pipeline Automation

Duration: March 2011 – August 2011

Tools used: MS Access, SQL Server 2008, MS Excel, McKinsey Periscope

Role: Data Analyst

Responsibilities:

- ✓ Data Ingestion
- ✓ SQL and VBA Scripting for ETL
- ✓ Flat file parsing
- ✓ Batch Scripting Data Pipeline Automation

INDUSTRY / WORK EXPERIENCE



Pratham Software Pvt. Ltd | Business Analytics & Solutions

Designation: Senior Data Specialist

Duration: November 2016 – Present



Aircel Ltd. | Business Analytics, Sales & Distribution

Designation: Manager

Duration: December 2015 – November 2016



Ernst & Young GSS Data Analytics - Advisory

Designation: Consultant/Senior Consultant

Duration: December 2014 – November 2015



Centurylink Technologies India Ltd. Marketing Analytics

Designation: Associate Analyst

Duration: September 2012 – December 2014



Bharti Airtel Services Africa Revenue Assurance

Designation: Senior Executive

Duration: August 2011 – September 2012



Evalueserve: Consumer Insights & Data Analytics

Designation: Business Analyst

Duration: November 2009 – August 2011

EDUCATIONAL QUALIFICATIONS



Panjab University

Bachelors of Engineering: Computer Science

Batch: 2005-2009

PUBLICATIONS

Science Fiction: Time Guard – The Awakening | ISBN: 978-1973354291