UDAYA CHITTA

Applied Data Scientist and Machine Learning Engineer
udaya.chitta@gmail.com
+91 939 172 6714
Bangalore, India

OBJECTIVE

Seeking a senior role as a Data Scientist for Fintech and Banking

SUMMARY

- Over 16 years of experience as an applied data scientist and machine learning engineer
- NLP Specialist
- Currently working as a modeler with Citibank, Personal Banking division
- I have built some critical data engineering assets for the bank backed both by relational data warehouses, SAS as well as big data platforms
- Conceptualized, Developed, Deployed and helping maintain a few ML, Deep learning and Statistical Models for Risk Scoring, Cross sell and a few merchant specific campaigns
- I have worked on a wide spectrum of products and services for the consumer banking *viz.* general purpose and store credit cards, deposits and investments, loyalty, personal loans, flex pay, brokerage and mortgage
- Did Consulting as a Data Scientist for 20+ Fortune 100 clients from Pharma, Hi-Tech, Manufacturing, Healthcare, OEM, Food and Beverage, Retail Consumer Goods and Automobile clients
- Multiple projects across various technologies - Python, R, Spark, Matlab, SAS and the various forms of SQL, Auto ML tools, Distributed computing, GPUs
- Take pride in coaching and mentoring - Have built and led mid sized teams for over ten years now.
- Seasoned in driving projects through their complete life cycle
- Experienced with both in premise and hosted cloud deployment of models and other solutions at scale
- Masters in Mathematics and Masters in Electrical Engineering from University of Illinois, Chicago
- Started my career Working in the US for 5 years for large banking corporations, enjoy a great deal of cultural compatibility working with diverse and global teams
- Pursuing research interests in the areas of Aspect Based Sentiment Analysis
- I apply a lot of emerging research and literature for my projects

EXPERIENCE

Citibank, Bangalore **August 2018 - Present**
Assistant Vice President, Personal Banking
Role -

- I am a part of the **Modeling, Data Innovation and Capabilities** team for the Global Consumer Banking
- I work on Data Powered innovation on customer spend, merchant analysis, loyalty, cross sell of personal wealth products among retail banking customers within US, Mexico, APAC and EMEA
- As part of this I Build new and improved data to solve specific problems in focus, combining the vast amount of internal data generated by product processors, warehouse data, data sourced from bureau, other external data such as market research databases, postal, census, government with customer demographic data
- I specialize in spend data and other transaction level data, customer journey data with an increasing focus on digital channels
- As part of this I Build new and improved data combining the vast amount of internal data generated by product processors, warehouse data, data sourced from bureau, other marketing databases
- Build Machine learning and Deep Learning Models
- Liaison with the business stakeholders and establish the modelling exercise
- Help wholesale banking with advisory and insights on the retail consumer spend

Capgemini, Bangalore **July 2013 - July 2018**
Manager, Insights and Data
Role -

- Build Open source Data Science Solutions in R and Python
- Build NLP solutions for predictive asset maintenance for Capgemini's various clients from IT, Manufacturing, Hi-tech, Retail CPG, Heavy Engineering, Energy and automotive industries
- Build domain centric data science solutions for Pharma clients in the areas of Key Opinion Leader identification and Information extraction to support Regulatory filings
- Build Big Data Engineering and Solution Architecture for High Performance Analytics over Spark, High Performance analytics on Massively Parallel Processing Data Bases, Migrate InMemory Python and R onto Distributed Environment

Quadrant Four Technologies, Chennai **June 2012 - July 2013**
Senior Data Scientist

- Role - Helped my client, a digital media company with automatic classification of their magazine content.

Tata Consultancy Services, Bangalore **July 2010 - June 2012**
Team Lead

- Role - Help my client, a Market Research firm, with statistical reporting for their Television Audience Measurement

Relocated to India, **June 2009 - July 2010**

Pyxis Solutions LLC, New York City **January 2005 - June 2009**
Consultant Business Analyst, Banking

RoleProgram the daily performance metrics for a hedge fund portfolio. Build a forecasting and reporting tool for Technology budget for a Bank.

PROJECTS

- **Complaint Identification from the bank's customer calls**
  A model for identifying complaints versus non complaints from among calls.
  Call transcripts are the source. Used a Pretrained Longformer for this.
  Got a 80%recall and a 70%precision
- **Call reason identification from the bank's customer calls**
  A multi class supervised model for identifying one among 20 top reasons for a call.
  This is a work in progress, using pretrained transformers.
  Also working on an unsupervised Aspect Based Sentiment Analysis model.
- **Predict Merchant Category from Google Search Results and Google Map Search Results** - Separated categorical features such as type of business from free text descriptions such as reviews and comments. Used tf-idf tokens from free text and one hot encoded categorical data into a XGBoost Model. With 7k observations across 800 plus business types, the model does considerably well with an 80%recall and a 60%Precision.
- **Automatic categorization of IT application maintenance incident tickets and Service Requests for Capgemini's clients**
  Mine the incident tickets and service requests from Enterprise Resource Planning software applications and Customer Relationship Management software applications running on a host of platforms such as SAP, SFDC, Oracle, BMC Remedy and Servicenow.
  These applications are designed, developed and maintained by Capgemini.
  NLP was done over ticket description text combined with application logs and system alerts
  Noun and Verb Phrases were used as tokens
  Support Vector Machine model was used over bag of words tf-idf feature set for multi label multi class supervised models.
  3 tier Labels were modeled for the three pillars - Problem description, Root cause and Resolution.
  The labels came out domain centric making the solution a great tool for automatic ticket routing and automatic ticket resolution.
- **A merchant recommendation engine for the bank's credit cards' customers**
  Used Singular Value Decomposition. Is aware of the customer's distance to the merchant location.
  Has a lift of 7%over the legacy collaborative recommender

- **Customer attrition model for the bank's deposits customers**
  Built a time to event model using survival analysis to predict the likelihood of attrition for a given customer at a future point in time.
  The model can also suggest the time period at which the attrition likelihood crosses a threshold.
  The model is built around different life stage and personal financial events leading to different customer states, attrition being one among those events.
  The state transitions are modeled as a Markov Chain. A competing risk survival model was used to predict the time to event.
- **A model for incremental response in Cross selling Deposit and Investment products to the bank's Credit Cards' customers**
  Used Zip+4 level affluence and property data from customer household level datasets from the bureau.
  Realized a 3%lift in response rate from top 2 deciles. XGBoost was used for the binary classification model.
- **Spend analysis**
  Helped more than 20 merchant partners with insights into Debit and Credit card spend volume, trend, correlated merchandise categories, triggered spend, spend across industry and competitor merchants, identify triggered spend and marketing opportunities.
  Merchants include Fortune 100 companies from Technology, Food and Beverages, Charity, Consumer Electronics, Retail, Pharmaceutical, and a few B2B companies as well.
  This was for the US and Mexico markets.
- **Predictive Asset Maintenance solutions from sensor data and other machine data**
  For Connected Cars, MRI devices, Network and Backup Servers. Sequence analysis, Natural Language Processing, Survival analysis are used.
- **Helped migrate Forensic Flight Data analytics from Database to Spark**
  Recoded the algorithms for Cumulative Damage Model, Flight Classification Models and migrated from InMemory R to pySpark. Also helped the ETL efforts from Oracle to Spark.