Manoranjan Kumar

Data Scientist (Manager) at Standard Chartered Bank – Global Business Services

I am an energetic professional with a successful record of embracing new challenges. I have 5+ years of experience of working on projects in the field of data science and analytics. In my current role, I contribute towards automation of finance data storage and reporting process and creating AI based financial insights. I come from a proficient academic background with a strong interest in problem solving and analytical thinking. I am looking forward to keep exploring the technological domain and apply it in different areas to solve real life challenges.

Education

2011-2016 **IIT Kharagpur**
M.Sc.(Int.), Department of Physics, 5 year course

2010 **ISC (Class XII)**
Xavier's English School

Work Experience

Jan 20 - Present **Manager (Data Strategy) I Standard Chartered Bank – Global Business Services**

At standard chartered, our team was built with the ambition of automating the financial and risk data storage and reporting process and then leveraging it in generating insights and modelling risk. In our day-to-day work, we analyze on financial risk data, commonly credit risk, market risk and operational risk and leverage machine learning techniques in solving the problems of banking and financial sector

Anomaly Detection in counterparty exposure data| Code: Python, DB: hdfs, hive
- Developed end to end data pipeline for this anomaly detection problem in the counterparty data
- Created an ensemble model of various time series anomaly detection models like one-class svm, Isolation Forest, Sarima, Inter Quantile Range method, Gaussian model etc. on this dataset

Entity Resolution for external counterparty | Code: Python
- Entity Resolution is the process of matching external sources data with that of internal records
- Performed text cleaning and transformations on these data and then applied several machine learning based models, like knn, fuzzywuzzy, hmni and word-tf-idf to do the name matching
- Currently it is productionized and extensively used by several teams in the organization

Early Alert Analysis| Code: Python, DB: hdfs, hive
- Created an Early Alert system for better assessment of risk and enhanced monitoring of portfolios
- Developed a Random Forest based classification model for different categorization of early alerts

Daily RWA (Risk Weighted Assets) Calculation tool | Code: Python, DB: Teradata
- Created python api for calculating Risk Weighted Assets on daily basis (earlier monthly basis)
- Used PD (Probability of Defaults), LGD (Loss Given Default) and EAD (Exposure at Default) data

LucidView(DQ Tool) | Code: Python
- Created a python-based tool for data quality, data validity and data completeness checks
- Tests include Null Check, Duplicate check, Outlier detection, Trend analysis, Descriptive Stats, Business rule check, range check, lookup check, pivot table analysis etc.
- various teams within the organization finds it very handy as it saves a lot of time and resources

March 19 - Dec -19 **Senior Data Scientist |Cornerstone OnDemand Pvt. Ltd.**

At Cornerstone OnDemand, I was part of R&D, Big-Data Machine Learning Platform Team. My position requires involvement in full development cycle of the product, **Idea-Brainstorming and feasibility study based on machine learning modelling/investigation in Python/R/SQL across clients -> UI specs design -> Backend Production Code Development -> Data validity, UI QA and Demo Story Preparation**.

Recommendation Engine for courses | Code: Java, Spark, Python
- This is a highly scalable recommendation engine in the LMS (Learning Management System) – HR space, currently empowering 3k clients using collaborative filtering and ALS (Alternating Least Squares). This aids employees in their learning path and career growth.
- Built the course and subject recommendation System in Python as a PoC and tested its accuracy
- Improved the model to recommend courses based on compliance/non-compliance course type filter
- Designed a model evaluation tracking system by measuring Mean Average precision (MAP) at K

Skills and Courses Data pool | Code: Python
- Created skills and courses data pool by crawling different sources
- Designed a word2vec based similarity matching process to map millions of skills and courses together
- Created a scorecard system to score a candidate resume matching his skills with the job description

Personal Info

Address :
565 A, Vijaya Bank Layout,
Bengaluru – 560076
Contact :
manoranjan.iitkgp@gmail.com
Phone : +91-7585965872
LinkedIn :
https://www.linkedin.com/in/manoranjan-kumar-38693249

Skills

Machine Learning
- Regression
- Classification
- Recommender Systems
- Time Series Analysis
- Clustering
- PCA

Deep Learning
- CNN
- ResNet, VGG
- Neural Style Transfer
- Siamese Model
- LSTM Encoder-Decoder
- Attention Mechanism

Backend Development

Software

Python	●●●●○
Java	●●●●○
Spark	●●●●○
Keras	●●●●○
Tensorflow	●●●○○
R	●●●●○
MongoDB	●●●○○
Hadoop&HBase	●●●○○
SQL	●●●●○
Dev-Ops	●●○○○

Relevant Courses

- Deep Learning.Ai Specialization (Coursera)
- Introduction to Statistical Learning (Book)
- The Analytics Edge (Edx)

Interests

Badminton
Technology
Social Science

July-2016 Feb-2019 **Data Scientist |Cornerstone OnDemand Pvt. Ltd.**

Survival and Recruiting insights | Code: Python

- Survival and recruiting insights serve many purposes. It is used to determine the right influx/outflux of employees in the organization and helps them in tracking this flow. Various insightful information can be deciphered from the Kaplan-Meyer survival curves, we plotted. It helps HR heads to find roles with high attrition and hiring latency to hire to be better prepared to either take preventive measures or fill them without much business losses.
- Did an extensive data analysis to figure out the key reasons for external attritions. It also helps an organization in managing their internal attrition carefully. Further, the recruitment funnel helps organizations in improving their recruitment process.
- Built a logistic regression classification model for predicting attrition risk of employees using java Spark MlLib, based on Learning data, Performance data & basic HR data.
- Achieved an average AUC of ~0.68 for several data-rich clients

Performance Rating Prediction | Code: Python

- An end-to-end data science exercise to predict the future performance ratings of employees using linear regression with R squared of 0.43, which is quite good in the social science community
- Results showed future performance depended heavily on predictors like past performance score, competencies scores, no. of competencies achieved, total years of experience, on-time completion rate and sometimes even the generosity of managers

May 2016 - July 2016 **Associate Data Analyst | Innoplexus Consulting Services Pune**

Classification of Pharma webpages into profile, forums, blogs, events & publications pages

- Worked on extracting data related to Pharma domain from online sources
- Performed data cleansing on the webpages data, removed stop words, did stemming and lemmatization and PoS-tagging
- Implemented BOW Tf-IDF SVM model for genre classification of webpages and research publications (text classification)

Personal Projects & Paper Implementation

March-2019 **Time series modeling for forecasting demand of electric installments in US across different subdivisions**

- Preprocessed the raw data to make it a time series object
- Did time series analysis to decompose it into trend, seasonality and randomness parts
- Plotted ACF&PACF graphs to get MA&AR components of the time-series data respectively and used Differencing to make it stationary
- Identified other variables which were impacting the dependent variable and tested the dependence using statistical tools
- Built an ARIMA model to forecast demand of electric installments across different subdivisions
- Determined the significant features by doing confusion matrix analysis and finalizing that recall is what the client should be optimizing
- Fitted an encoder-decoder LSTM model as well for this time series forecasting and got an improved accuracy

June-2019 **Paper implementation of the ResNet paper [He et al., 2015. Deep residual networks for image recognition] (https://github.com/manoranjan03/PaperImplementations.gi**

- Trained a 20-layer, a 32 layer, a 44 layer and a 56 layer plain network as well as the corresponding residual networks on the famous cifar-10 and ImageNet dataset and plotted their validation and train error for multiple epochs
- Verified that the training error (and not only the validation error) increases as the network becomes deeper for plain networks and hence overfitting is not the cause of this accuracy degradation
- Proved that contrary to the plain nets the training and val errors decreases for ResNets with an increase in the depth of the network
- Compared the validation data accuracy of different models (resnet-20, resnet-32, resnet-56, plain net-20, plain net 32, VGG-16) and proved that using the residual blocks not only increases the accuracy, but it also reduces the model complexity and computation
- Verified that the overfitting kicks in residual networks as well if it is too deep (over 1000-layers)

Publication & Research Project

Jul 2014 - May-2016 **Ultra Fast Science, Department of Atomic Energy(DAE) and IIT Kharagpur**

Dr. P. K. Datta, Manoranjan Kumar, “Investigation on the stability of cascaded second order Mode-locked laser ”, in the proceedings of the conference “Ultra-Fast Science-2014”, a DAE – BRNS Theme meeting on Ultrafast Science – Mangalore

- Modelled a stable femtosecond pulsed laser system that would increase the accuracy of medical diagnoses and treatment by 1000 times
- Modulated and solved Master Mode Locking Equation for the experimental setup of cascaded second order Mode- locked laser
- Analyzed the experimental data using simulations in MATLAB to derive different parameters to prove the increase in gain by 10 times
- Derived the stability condition theoretically and used it to prove the improved stability of the Mode Locked laser(Nd:YVO4;) by 100 times
- Measured the non-linearity and proved the multi-photon absorption processes by simulating the z-scan experimental data in MATLAB
- Did a poster presentation of our research work in the conference UFS-14

Academic Pursuit

Aug2019- May2021 **Master’s in Business Law (MBL) | NLSIU (National Law College, Bangalore)**

Contract Law | Corporate Law | Intellectual Property Law | Investment Law | Banking Law

- MBL is a two-year post-graduate degree course with a focus on applying concepts of law to business domain. The topics covered in this course come from the diverse domain from Bankruptcy law to Intellectual Property Law, Our Corporate Law , and Investment Law and is more oriented towards solving practical problems.
- The idea is to understand the evolution of legal systems across time and geograpies and its impact on our current society and how proper legal encoding can be helpful in reshaping our future.