## AAKASH BASNET

### SUMMARY

* 2+ years of experience in **Artificial Intelligence**, **Machine Learning, Statistical Modelling, Data Science, Data Mining, Algorithms and Data Structures**  and **Web development**.
* Proficient in **machine learning** and **deep learning** skills for multiple application including **Computer Vision**, **Recommendation Systems** and **Natural Language Processing**.
* Experience with mathematical and statistical Python Libraries such as **pandas**, **skit-learn**, **NumPy**, **NLTK**, **Pytorch**, **SciPy**, **TensorFLow**, **Keras** and software such as **MATLAB** and **R**.
* Deployed, debugged and maintained complex, distributed software stacks, containing **Apache Spark**, **Hadoop HDFS** and **IPython Notebook** servers, on cloud-based **AWS** system by optimizing the stacks for best computational performance and stability.
* Extensive experience with **advanced mathematics**, **statistics**, **applied machine learning** and **visualizing** complex data and concepts to diverse audiences.
* Expertise in transforming business requirements into **analytical models**, **designing algorithms**, **building models**, developing **data mining** and reporting solutions that scales across massive volume of structured and unstructured data.
* Experienced in generator functions and generator expressions.
* Hands on experience in using **Tensorflow and Probabilistic Graphical methods(Bayesian and Gaussian networks)** to create machine learning models.
* Ability to document ML project requirements and assess deliverable timelines.
* Excellent working knowledge in **UNIX** and **Linux** shell environments using command line utilities.
* Experienced with **GPU computing** and **data mining**.
* Experienced using tools for **MDM(Master Data Management)** like Oracle MDM and Reltio
* Experienced in implementing data analytics solutions in **AWS (EC2, EMR. S3), Google Cloud, Microsoft Azure**.
* Used **AWS Lambda Functions** to trigger the code and managed custom performance and security along.
* Experienced in **Agile**, **Scrum, Waterfall** and **Sprints methodologies**.
* Experienced with **object-oriented programming** (**OOPS**) concepts using **Python**, **C++ and Java**.
* Excellent skills in analyzing big and complex data, data matching, identify anomalies , and providing usable insight to internal and external data activities and trend
* Proficiency in implementing **Multi-threaded** applications and using **Super computer servers**
* **Web/Application Development** with deep understanding of the technologies which involves analysis, testing, design, development, implementation and maintenance of various web based application using **Python** and **Django**.3+
* Experienced in developing consumer-based custom features and application development using **Python**, **Django**, **HTML5/CSS**, **XML/JSON**, **JavaScript**.
* Vigorous knowledge in progressing **web applications** and effectuating **Model View Control** architecture using **Django** web application framework.
* Good knowledge of web services with **protocols SOAP, REST**.
* Strong Communication and Presentation Skills substantiated in past assignments with developers, project managers, subject-matter experts, stakeholders, system implementers, and application end-users.
* Experienced with **Unit Testing**, **System Integration Testing** (**SIT**) and **User Acceptance Testing** (**UAT**).
* Good experience in **error** and **exceptional** handling.
* Good knowledge of NoSQL databases like MongoDB and Cassandra and as well as relational databases.
* Experienced with version control systems like **Git**, **SVN** to keep the versions and configurations of the code organized.
* Hands-on experience in **writing** and **reviewing requirements, architecture documents, test plans, design documents, quality analysis and audits.**
* Good knowledge on **Database Definition Language**, **Database Design**, **Data Warehouse Design**.
* Experienced in designing the automation framework using **Shell scripting**.
* Proficient in **research** of current process and emerging technologies which need analytic models, data inputs and outputs, analytic metrics and user interface needs.
* Quick learner and keen to adopt the cutting-edge technologies .

### 

### SKILLS

**Operating Systems:** Linux, Windows, Mac

**Programming:** Python, Java, MATLAB, C, C++,C#, R, Haskell, JavaScript, HTML, CSS, Latex

**Databases:** SQL, PostgreSQL, MySQL, Oracle, NoSQL, MS Access, Hadoop(for big data)

**Version Control:** Git, SVN

**Cloud Services:** AWS, Azure, Google Cloud

**Data Analytics/visualization tool:** R, Microsoft Power BI,Microstrategy, Tableau, d3.js Pentaho, STATA, NVivo, Informatica, Hadoop, Accumulo (Google big table concept), HBase, Spark, Qlick, Flume, MQ Services, Sqoop, Elastic Search, MapReduce, Amazon S3, Azure, Zeppelin, Yarn, Hive, Python IDE

**Machine Learning/AI:** Artificial Neural Network, Convolution Neural Network Bayesian Network/BBN, Linear Regression, Logistic Regression, Decisions Tree, Random forest, Pruning, k-NN, SVM, SVDK Clustering, Page Rank and PCA, MCA, MFC, Apriori and other data mining and Deep Learning Algorithms, Social Media Analytics, Sentimental analysis, Market Basket Analysis, Bagging, Boosting, Reinforcement Learning, Q Learning, Hidden Markov Models, Feature selection.

**Tools:**  Visual Studio, IntelliJ, Pycharm, Android Studio, Putty, Filezilla, TFS, JIRA, Rally, Version1, HP ALM, Test Track Pro, Rational team Concert

**Methodologies**: Agile, Scrum, Waterfall

### 

### WORK EXPERIENCE

**UNMBBER, Albuquerque, NM May 2018 - Dec 2018**

**Python Developer/ Machine Learning**

The UNM Bureau of Business and Economic Research (UNMBBER) is the recognized expert in providing socioeconomic data and forecasting in New Mexico. UNMBBER’s research team provides economic forecasting as well as economic research services and data communication tools tailored to the needs of clients – public, private, nonprofit, and philanthropic – seeking to understand and shape public policy on the state, regional, and local levels. UNMBBER’s services and research help leaders in New Mexico to understand, forecast and identify trends and changing economic markets across the state of New Mexico in order to inform decision making.

**Roles and Responsibilities**

* System administrator for servers on Amazon Web Services: Handling production and staging server instances, Server Database Management, Creating API node for new functionality
* Managed database of more than 10MM records of data by writing scripts and macros to automate data updating and cleaning process which saved monotonous working hours.
* Responsible for architecting, designing, implementing and supporting of cloud based infrastructure and its solutions.
* Involved in writing Python API for Amazon Lambda to manage some of the AWS services.
* Performed data pre-processing to clean, eliminate outliers in data and conducted data exploration to detect correlation, trends and patterns in the data
* Imported the big socio-economic data files , created functions to read and join the files and generated data visualizations of state wise statistics of the data using Python libraries and d3.js for displaying in the web and connected all the visualization directly to the database
* Utilized PostgreSQL , data warehousing programs, Tableau, and other dashboard/visualization tool sets for data intelligence and analysis.
* Linked all the critical data to a common point of reference using Reltio(MDM) and sharing among different group of researcher.
* Designing, reviewing, implementing and optimizing data transformation processes in the Hadoop and Informatica ecosystems
* Led several big data machine learning initiatives involving the design, development and deployment of advanced machine learning algorithms that impacted more than 1000 local business.
* Predicted how sick leave would affect the local and small scale business in New Mexico in upcoming 10 years, by leading the survey team to collect and clean the data and preparing the Machine learning model.
* Worked with the research team to forecast the poverty rate of New Mexico in upcoming years by building the machine learning model for the large set of Census data of New Mexico.
* Build the ML model to forecast how much a new company, that is being established, would increase the employment rate and GDP in the present economy.
* Build sentiment analysis model using Natural Language Processing (using NLTK) for new startup in New Mexico are based on customer reviews.
* Worked with several government client, like MRCOG ( Mid-Region Council of Governments) by providing them with clean data set and visualization for regional development planning. Build the water level visualization for all the water resources in New Mexico by collecting the data from the on-site sensors that would update in every 5 minutes.
* Worked parallely with the research team to forecast the impact of new start up in the economy of New Mexico.

**DataRobot, New York, NY May 2017 - May 2018**

**Customer Facing Data Scientist**

DataRobot is an Automated Machine Learning product firm. Helped DataRobot acquire new clients working with decision makers, conducting Proof of Concepts, nurturing prospects into clients and developing collaborative culture.

**Roles and Responsibilities**

* Ensured alignment with key technology and business stakeholders across globally diverse, Agile teams.
* Helped Bank of America through building predictive models in FX, Fixed Income, Investment Banking, Research, Commercial Lending, Wholesale Credit, Client Relations, M&A.
* FX Volume Prediction: built time series models to predict daily FX volume for CLS at a minute level using data from primary exchanges - EBS, Reuters, BofA volume, bid and ask rates, spreads, VWAPs, simple and exponential moving averages, order book entries, etc. Used Bollinger Bands, MACD, market events, holidays for EUR/USD, USD/CAD, etc currency pairs. MASE values were impressive compared to a naïve model.
* Customer Attrition Prediction: built a highly successful customer attrition predictive model with 80% accuracy on FICC electronic trading from Bloomberg terminals using time series, feature engineering with financial ratios, etc.
* Capital Review Committee Revenue Prediction: predicted yearly revenues for years 1 to 3 for the bank on 16 products ranging from Treasury, Advisory, Credit to FX, and had beat bankers estimates.
* Funded Loan Growth Prediction: developed predictive models for funded loan growth for the Corporate Banking group at Bank of America and improved prediction probability six times. Found key drivers and early indicators.
* Worked with Balyasny Asset Management (BAM) hedge fund, JPMC, TD Bank, GRA (Global Risk) at BAML
* Used Oracle MDM to share critical data among different departments and personal.
* Built custom machine learning models on large datasets in use cases such as optimal capital allocation, commercial loan growth, customer attrition, market trend prediction for the bank.
* Built multi-class sentiment analysis models on bank’s research reports using NLP and Spacy.
* Moved Machine Learning projects into production and created tangible value for the firms.
* Brought business insights showing feature interactions in ratings tables, prediction explanations.
* Built several workflows that combined data preprocessing steps with feature engineering, feature selection, model selection, hyper-parameter tuning, model stacking, blending, using cross validation to avoid overfitting, validating models with lift charts and ROC curves, explaining insights through feature importance analysis, partial dependency plots. Handled class imbalance and large datasets. Explored human – machine hybrid approaches.
* Captured trends, seasonality patterns through time series models such as ARIMA, used lag variables and sliding window techniques, feature engineered variables through iterations.
* Analyzed unstructured text in analyst reports, built sentiment analysis using TFIDF, NLP, Spacy.
* Balanced algorithm accuracy over speed in XGBoost, Random Forest, GLM, ENet Blender, Logistic.
* Worked with bank regulators on variable stress testing, parameter sensitivity analysis.
* Built challenger models for BAML regulators, the Model Review Management group, a three month long process, with variable stress testing, hyper-parameter sensitivity analysis, out of time validation, and model deployment.
* Helped Humana insurance with Marketing mix optimization, Emergency Room attendance estimates.
* Developed Oil recovery models for Hunt Oil, and transport ETA predictions for Rail Inc. and BASF.
* Handled large scale transactional, trading, loan, hospital, transportation, oil production data.
* Evangelized Artificial Intelligence, Machine Learning through presentations, online webinars, blogs.
* Wrote popular blogs on Machine Learning and received company’s special award on content creation.

**Environment:** Python, R, SKLearn, Time Series, ARIMA, Multiclass, Anomaly Detection, Feature Engineering, Imbalanced data, SQL, Hive, Hadoop, Tableau, Spacy, NLP, Spark, DataRobot, Eureqa, Nutonian

### PERSONAL PROJECTS **Kaggle:**

https://www.kaggle.com/hssaka7

Human Protein Atlas Image Classification

* Was provided with 250 gb (3072 x 3072 TIFF) image files from Human protein Atlas dataset.
* Predicting protein organelle localization labels for each sample. There were in total 28 different labels present in the dataset. The dataset was acquired in a highly standardized way using one imaging modality (confocal microscopy). However, the dataset comprises 27 different cell types of highly different morphology, which affect the protein patterns of the different organelles. All image samples were represented by four filters (stored as individual files), the protein of interest (green) plus three cellular landmarks: nucleus (blue), microtubules (red), endoplasmic reticulum (yellow).
* Preprocessed the given data and used Convolution Neural Network to predict the class for given protein image.
* Used GPU and server computing on AWS to run the code, since the data were too big and would take much time to pre-process data and fit the model.

Music Genre Classifier

* Used the GTZAN dataset, which is frequently used to benchmark music genre classification tasks. It was organized into 10 distinct genres: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. The dataset contains the first 30 seconds of 100 songs per genre. The tracks were recorded at 22,050 Hz (22,050 readings per second) mono in the au format.
* Classified music files into 10 predetermined genres such as jazz, classical, country, pop, rock, and metal by building and testing different machine learning models.

Topic Categorization

* Dataset was obtained from 20Newsgroups. The 20 Newsgroups dataset is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups. This collection has become a popular dataset for experiments in text applications of machine learning techniques, such as text classification and text clustering.
* Used Bag of Words model, (writing Naive Bayes and Logistic Regression from scratch) for building Machine Learning algorithm to predict the category of the given news group.

**Other Projects :**

NASA Swarm Base Code

* Designed and conducted experiments on Swarm Robots and Simulated code through aRGOSand ROS environment.
* Deployed the AI code on real robots and prepared the experiment reports.

PacMan Agent

* Implemented different AI algorithms like A\* , Reinforcement learning, Q learning on PacMan Game agent to solve the given maze.
* Trained the maze using different Machine learning algorithms and feature selections to solve the most difficult maze available.

More projects on : https://github.com/hssaka7

### EDUCATION

Bsc in **Computer Science** University of New Mexico (School of Engineering)

with a minor in **Math** GPA: 3.6

### VOLUNTEER

**Center for Academic Program Support** | STEM Tutor

* Provided academic support to over 2,400 unique students each semester and helping to develop and expand computer science, mathematics (college algebra - upper label math), University physics and programming skills.
* Led 4/12 weekly mandatory training each semester to enhance better tutoring skills and design tutoring related materials and blogs.
* Volunteered in creating the student database and tutoring related software/materials like projectile calculator, blogs and videos .