ARPIT KULSHRESTHA

Technology Architect, 87-801 Jameson Ave, Toronto, M6K 2W8
+1 416 876 3727
https://www.linkedin.com/in/speak2arpit/
https://github.com/Arpit1989

EDUCATION

- Currently pursuing Master of Data Analytics with concentration in applied machine learning from Northeastern University, Toronto, Canada, 2021-2023 (Current GPA : 3.852/4.0)
- Bachelor of Technology from SRM University, Chennai, Tamil Nadu, India, 2007- 2011 (CGPA: 7.6/10)

EXPERIENCE SUMMARY

- +10 years of experience in Java, Javascript, Ruby, Golang, Python, R.
- Extensively worked on frameworks like Springs, Hibernate, RubyOnRails, AngularJs, ReactJs, ExpressJs, NextJs, NodeJs.
- Expert with search engine, worked on Solr, Elastic search, IBM Vivisimo, semantic search.
- Experience in working on Redis, Cassandra, SolrJ, dbpedia, Apache Jena, Apache Tika, Semantic web, Ontologies, Data Mining, Text Analysis and Extraction.
- Experience in Agile Methodology from Sprint planning, Poker Planning, Story estimation Techniques to Burn down/up charts.
- Won 700$ in India Solar Hackathon organized by DevPost in India for an AI Bot using dbpedia (2016, https://devpost.com/software/solarkohli).
- Created an AI bot using Python, and Flask which provides the Indian Railway Information for current status of the running trains and PNR enquiries.
- RubyConf India Goa 2014 - lightening talk on linked data with ruby.
- Experience in enterprise application life cycle design, development and deployment.
- Expertise in Data analysis; regression/classification/cluster analysis, algorithm comparison, building machine learning model using TensorFlow.
- Data visualization via different plotting libraries like D3, ggplot, plotly, maps etc.

Programming Experience

OPERATING SYSTEM	PROGRAMMING LANGUAGES	FRAMEWORKS & LIBRARIES
UNIX based, Windows, Android, iOS.	Java, Ruby, Python, TypeScript, Javascript, Swift, Applescript, Golang, Scala, SQL, SPARQL, R.	ReactJS, NextJs, NodeJs, AngularJs, TensorFlow, GoogleAutoML, Apache Tika, Apache Jena, Tomcat, RubyOnRails, Springs, Hibernate, WebSockets.io, Solr, Elastic Search, IBM vivisimo , Facebook APIs, Soundcloud APIs, Triple Stores, Docker, Jenkins, Rspec, Cucumber, dplyr, ANOVA, Regression, classification, RShiny, RShinyDashboard, Tableau, StrapiCMS.

NORTHEASTERN UNIVERSITY - AUG 2021 - Current

Project Title	Toronto DineSafe Program
Period	Mar 2022 - Jul 2022
Position	Student of Analytics at Northeastern University
Responsibilities	Analyzed the Toronto DineSafe program data feed shared at Open data portal Toronto.
Project	Classify the restaurants based on geography, infractions, and violations. Conclusion are listed below. • There were no visible patterns of infractions that can be geographically distinguished from one another. The inspection is done at random and infractions are not geographically concentrated in any certain region. • The Child-Care Catering and Child Care Food Preparation has the least number of infraction per restaurants. • The mobile food preparation has the worst ratio of infraction/type which is understandable as these are mobile food preparation Units, requires high maintenance compared to other types of services
Hardware	MacOS
Languages	Python
Database	XML Feed
Tools & libraries	jupyter notebook, numpy, pandas, ElementTree, plotly, sklearn, json.

Project Title	Merry Maids Data analysis
Period	Jan 2022 - Jul 2022
Position	Student of Analytics at Northeastern University

Responsibilities	Build, design, analyze and test contact prioritization model for Merry Maids brand. The project is divided into multiple sections that are; the research methodology used to derive the models, exploratory data analysis to understand each parameter, feature selection and importance, the results of each machine learning model, and how these models can be utilized for contact prioritization. The objective is to ensure that the sales team contacts prospects with a higher likelihood of conversion. Our results demonstrate that a machine learning model can be utilized for contact prioritization and can outperform the existing method of not prioritizing contacts.
Project	Lead Contact Prioritization Model for Customer Acquisition at Merry Maids Using Machine learning algorithms.
Hardware	MacOS
Languages	R language
Database	RDS
Tools & libraries	janitor, ggplot, caret, tidyr, dplyr, purrr, broom, mice, corrogram, and many more.

Project Title	Toronto Regional Real Estate Board Data Analysis
Period	Jan 2022 - Mar 2022
Position	Student of Analytics at Northeastern University
Responsibilities	An analysis of Toronto Real estate prices on data published quarterly by the TREB. MLS Score indicator and price comparison, most and least affordable housing options based on area.
Project	The project aims to build a tableau dashboard that can answer different business questions based on the real estate prices. Some of the findings are listed below: • Did you know that the average price of real estate properties have risen by at least 70% in the last 5 years in Toronto 2017-2021. • Did you know apartments are the most affordable housing type in Toronto. An average price around 500K$. • Toronto eastern district 01 is the most expensive area to live in
Hardware	MacOS
Languages	R language

Database	XLS
Tools & libraries	Tableau dashboard, RStudio, ggplot, dplyr, tidyr, EDA.

Project Title	Ontario Covid Data analysis
Period	Sept 2021 - Dec 2021
Position	Student at Northeastern University
Responsibilities	The primary objective of the report is to answer the questions related to covid-19 pandemic via statistical methods like ANNOVA, ChiSq Test, Regression Analysis & etc. The report also includes trends, patterns identified during the analysis and a forecast model based on ARIMA which combines unit root tests, minimization of the AIC and MLE to obtain an ARIMA model.
Project	The project aims to answer these specific business questions and trends of covid-19: 1. Is there any significant difference in the number of people died in each of the covid-19 wave? The daily number of deaths remains the same through different waves. 2. Is the outcome of the covid-19 infection dependent on the medium/mode of spread? We want to find out if a result of covid infection (fatal and non fatal) is dependent on the mode of spread (was the infection due to close contact, community spread, or travel or outbreak related) 3. Is the outcome of the covid-19 infection dependent on gender of infected individual? We want to find out if a result of covid infection (fatal and non fatal) is dependent on the gender of infected (are male more susceptible to fatal outcome?)
Hardware	RStudio MacOS.
Languages	R language
Database	None
Tools & libraries	RStudio, ChiSq Test for independence, AR/MA/ARIMA.

Project Title	Job Search Engine keyword analysis
Period	Sept 2021 - Dec 2021

Position	Student at Northeastern University
Responsibilities	Data scraped from LinkedIn search, Indeed search for different keywords and performed an data exploration
Project	Project aims to understand the current job market in Canada for data science/analytics. It is an exploratory data analysis that requires scrapping data from job portals.
Hardware	RStudio MacOS.
Languages	R language, Python
Database	None
Tools & libraries	BeautifulSoap, pandas, RStudio, ggplot, dplyr, tidyr. GLM

FREELANCE 2021-2022

Project Title	Wsh Future
Period	Jun 2022 - Ongoing
Position	Technology Architect/ Cloud Solution Architect/Developer
Responsibilities	Build, design, architect an investment platform for investors and startups.
Project	The project aims to provide a platform similar to seedrs, republic for Indian market.
Hardware	AWS cloud services
Languages	Typescript, Ruby, Javascript
Database	Postgres
Tools & libraries	S3, EC2, RDS, SES, Strapi CMS, Ruby on rails, GraphQL, SEO, User Analytics.

Project Title	Oto lawn cloud services
Period	Mar 2022 - July 2022
Position	Technology Architect/ Cloud Solution Architect/Developer

Responsibilities	Migrate the existing cloud services from cloud functions to cloud run, create CI/CD framework for deployment from Azure Devops Pipeline to Google Cloud using Artifact repository. Design the new authentication mechanism and migrate existing users from firebase.
Project	Awarded as the Best Innovation of 2022 at CES 2022, Oto Lawn is a complete lawn care solution that comes with an IoT device. https://otolawn.com/
Hardware	AzureDevOps, Google Cloud
Languages	Java, Javascript, Shell Script
Database	MongoDb
Tools & libraries	Spring Boot, Spring, JPA, React, AzureDevops Pipeline, Google Cloud Run, Artifact Repository,

INFOSYS 2019-2021

Project Title	Patient Care Central with CVS
Period	Apr 2020 to May 2021
Position	Technology Architect
Responsibilities	Build, lead, review, and design microservice for the patient profile, store, prescription, orchestration, user interface, and work closely with the senior project architects to design solutions. As part of modernization and digitalization, we are replacing a monolithic legacy application used by pharmacists with cloud-hosted microservices and a modern User interface.
Project	Patient Care Central is an application built to extract features from a legacy product used by the store physician at CVS. In phase 1, Aim is to extract opportunities and patient reach. The project is implemented using Pivotal Cloud Foundry to host the microservices, Jenkins as CI-CD.
Hardware	Pivotal Cloud Foundry
Languages	Java, Angular 7
Database	Oracle
Tools & libraries	Spring Boot, Spring, JPA, Hibernate, Hysterix, Informatica.

INFOSYS 2019-2020

Project Title	Warehouse Management System at Macy's
Period	March 2019 to Apr 2020
Position	Tech Lead/Team Lead/Full Stack Developer
Responsibilities	Design, plan, and review warehouse management microservices, specifically in the planning space. Led a team that created waving and distribution solutions from scratch. Worked on services including wave, distribution engine, order fulfillment services, handheld service, supply chain UI, handheld UI, and work closely with the project architects on designing solutions. Worked closely with a data scientist on optimization using a gurobi engine for the waving process.
Project	Backstage WMS consists of microservices like – Inventory, Work State Management, Order Management, Waving, Distribution Engine, Handheld; all of these services were used to run the warehouse, which was using Pyramid as WCS (Warehouse Control System) and WES (Warehouse Execution System). The project was implemented using Google Cloud to host the microservices, Google Pub-Sub as messaging service, Google Data flow jobs for auditing using BQ, Jenkins as CI-CD, Kubernetes for Automated deployment, and scaling.
Hardware	Google Cloud Platform
Languages	Java, React
Database	BigQuery, MySql,Redis
Tools & libraries	Spring Boot, Spring, JPA, Hibernate, Sleuth, Prometheus, Spring Contracts, Redission, Liquibase, Gurobi Engine, Manhatten WMS.

FREELANCE DEVELOPER 2018-2019

Project Title	Image Classification
Period	Oct 2018 - March 2019
Position	Collaborator/Advisor
Responsibilities	Evaluate available image classification solutions, including GoogleAutoML, Amazon Rekonginition, IBM Watson, tensor flow. Create a dataset with the desired label for training purposes. Upload all the files to google cloud storage and create an AutoML project with mapping and label information. Train the model for the desired number of hours, check the accuracy for each label with the recall rate. Repeat the process until we get the desired accuracy.

Project	Classification of images collected by parks, recreation, and tourism department using Machine learning via Google AutoML; as part of the project, we started with creating our modal using tensor flow and then moved on to Google AutoML Vision, which uses tensor flow light to create a machine learning model. The goal of the project was to create a solution to classify, count objects, and tag their images collected from various national parks using Google AutoML and YOLO faster RCNN. The images of previous years were used as a training data set.
Hardware	Google Cloud
Languages	Python
Database	Cloud Storage
Tools & libraries	AutoML, Tensor Flow, FigureEight, Compute instances, MsExcel, yolo.

FREELANCE DEVELOPER 2018-2019

Project Title	3xFive
Period	Aug 2018 - March 2019
Position	Software Architect/Advisor
Responsibilities	Create a product of funnel management for Inquire Media inc. Design, Architect a cost-effective solution and scale to handle thousands of requests per second—building the entire UI framework using NextJs, customizing for each funnel.
Project	The funnel management platform creates static web pages using a customized editor to choose the color, fonts, images, and form styles. Advertisers use these hosted funnels can route traffic to these static pages that request back-end services.
Hardware	AWS
Languages	Javascript, Ruby
Database	Amazon RDS, Amazon S3
Tools & libraries	Amazon EC2, RubyOnRails, NextJs, ReactJs, bandwidth.

MAROPOST 2017-2018

Project Title	Maropost Marketing Cloud
Period	Oct 2017 to Aug 2018
Position	Technical Project Manager

Responsibilities	Review, manage, gather requirement for the Maropost Marketing Cloud Development, Worked on POCs to improve the Maropost Marketing Cloud while also handling the review, architecture of the new features, and its impact on the marketing cloud. Responsible for keeping the VicePresident and the Board Members up to date on the development of Maropost Marketing Cloud features. As part of the POC's, we introduced GrapeJs into marketing Cloud, ElasticSearch, to improve data discoverability. As a team, we architected the Social Media Marketing Cloud. MMC integration with Zapier, Created a Magento Plugin for real-time syncing of orders Improved the user tracking script to work on all modern-day web platforms. Helped team to create and build an Acquisition platform for marketers to acquire new leads for clients.
Project	Maropost Marketing Cloud - It was a platform built for Marketers to help them send campaigns to their customers through different channels such as Email, SMS, Push messaging.
Hardware	Google Cloud Platform and Rackspace Unix Machine.
Languages	Ruby, Go, Scala.
Database	MySQL, CASSANDRA,POSTGRES,
Tools & libraries	RubyOnRails, Apache Spark, Apache Kafka, ReactJs, VueJs, WebCrawler, User-Tracking through In-house Tool, ElasticSearch, Kibana, FreshDesk, ATOM, Pivotal Tracker.

Project Title	Maropost Marketing Cloud and Sales Cloud
Period	Aug 2017 to Oct 2017
Position	Senior Developer
Responsibilities	Introduced Elastic Search, Kibana for reporting and querying, evaluated features of Salesforce, and built a framework for similar solution. Worked with the CTO in designing the Sales cloud. Responsible for handling and architecting the solutions for the Maropost Marketing Cloud and its features. My role started from developing some of the features to then guiding the team of over 40 developers to deliver the solution. Worked with the Vice President of Maropost to help them deliver quality features such as Social Media Marketing Platform for Facebook, LinkedIn, Twitter.

Project	Maropost Marketing Cloud - It was a platform built for Marketers to help them send campaigns to their customers through different channels such as Email, SMS, Push messaging, Social Media Marketing. It integrates well with several tools like salesforce, zapier, Magento, Shopify and many others. Maropost SalesCloud - A platform which we were building offered CRM, e-Commerce, Customer Support Ticketing, Referral Management. Vision was to create an ecosystem of these multiple application which can share data among them.
Hardware	Google Cloud Platform and Rackspace Unix Machine.
Languages	Ruby, Golang, Scala.
Database	MySQL, CASSANDRA, POSTGRES,
Tools & libraries	RubyOnRails, Apache Spark, Apache Kafka, ReactJs, VueJs, WebCrawler, User-Tracking through In-house Tool, ElasticSearch, Kibana, FreshDesk, ATOM, Pivotal Tracker.

TATA CONSULTANCY SERVICES: 2011-2017

Project Title	Classification Tool
Period	June 2015 to July 2017
Position	Lead Developer & Ontology Expert for Mckinsey & Company.
Responsibilities	The project's goal was to research and evaluate the platforms available for replacing an old taxonomy tool that used the traditional relational database to manage topics and sub-topics for classification of documents within McKinsey & Company. We evaluated several open source and commercial solutions based on different parameters, which later got published internally to build the ontology management solution centered on our recommendations, Data mining, Requirement gathering and analysis, Estimation of the software development effort, Design and Development of the application, Unit Testing and Defect Verification, Application Deployment. After the Evaluation, the project's goal was to replace an old taxonomy tool, which was based on the relational database. As part of the setup, we also enhance the existing taxonomy by mining other linked databases.
Project	Ontology Management platform which allows users to suggest, add, remove ontologies within the organization. Springs Application which exposes APIs and Interact with and Angular JS Application. This Application uses NoSQL database using Apache Jena and Interacts with Oracle Database. Primary Purpose of the application is to manage the knowledge of the organization using ontology.
Hardware	Windows 12 with 8 core 2.8 GHz Intel Core i7 processor, 30 GB RAM
Languages	Java
Database	Triple Store and Oracle 12

Tools & libraries	Spring Tool Suite, Solr, PoolParty, Hibernate,Apache tika, RDFs, Linked Data, Spring security, LDAP, SSO, Junit, Mockito, IBM Lotus Notes, Swagger, SKOS.

Project Title	Know
Period	Jan 2014 to June 2015
Position	Java Developer and Search expert for McKinsey & Company
Responsibilities	As an agile developer perform the following based on story boarding- Solr Configurations, document extractions, requirement gathering and analysis, estimation of the software development effort, design and development of the application, unit testing and defect verification, application deployment.
Project	To provide a search platform for searching all the internal documents using Solr. The data was imported to Solr using DIH, and Incremental updates were performed using SolrJ. Java web application provided the interface to query the Solr Data for the user to search. The entire Solr Cloud was configured as part of the project, which involved using various filters and stemmers. Content extraction using Apache Tika & indexing those extracted text based on the language and filters.
Hardware	Suse server 4GB RAM
Languages	Java
Database	Oracle
Tools & Library	Solr, SolrJ, JSP, Javascript , Jquery , Html5, Css3, Apache Tika,

Project Title	Data analysis and semantic search
Period	Jan 2013 to Dec 2013
Position	Java Developer, Linked data expert for McKinsey & Company
Responsibilities	As an agile developer perform the following based on story boarding- requirement gathering and analysis, estimation of the software development effort, design and development of the application, unit testing and defect verification, application deployment.
Project	The project involved data mining the open source linked databases for improving the current taxonomy of data from industry, functions to geography. The project used geo names, dbpedia and other open source databases to mine the relevant data.
Hardware	MacBook Pro with 2.6 GHz dual-core Intel Core i3 processor, 4 GB RAM
Languages	Java, XML
Database	Triple Store

Tools & libraries	Apache Jena, Solr, Dbpedia, Linked data, Apache Tika

Project Title	Alumni Center
Period	Dec 2012 to Jan 2013
Position	Ruby on Rails Developer for McKinsey & Company
Responsibilities	As an agile developer perform the following- requirement gathering and analysis, estimation of the software development effort, design and development of the application, unit testing and defect verification, application deployment.
Project	The projects involved upgrading the ruby and rails version, features for searching alumni and its outreach, site core integration, new ui designs.
Hardware	MacBook Pro with 2.6 GHz dual-core Intel Core i3 processor, 4 GB RAM
Languages	Ruby
Database	Oracle
Tools & libraries	Ruby on Rails, Solr, SiteCore CMS, CDN.

*Please refer my github account for more details on the projects, Over 10K download of the libraries I have published/contributed in ruby gems.