Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

Executive Summary:

- Working in **Multiple Data Scientist, Data Analyst, Data Engineer Projects** Includes Experts in **Artificial Intelligence, Machine Learning**, Deep Learning, RPA, ETL (Informatica Power Center) DWH, BI, Data Visualization (Tableau), OORDBMS (SQL, PLSQL, Toad for Oracle, SQL*Loader), Bigdata (Hadoop, HDFS Architecture, Map reduce, Spark, Scala, RDD, Sqoop).
- Experience in **Anaconda Jupyter Notebook, PYTHON Programming**, Python Libraries (NumPy, SciPy, Seaborn, Pandas, SK Learn, Matplotlib + PyPlot) and Statistical Application in R- Programming.
- Strong in **AI Tools & Frameworks** (Tensor Flow, Theano, Keras, PyTorch, CNTK, MxNet, Caffe, Auto ML, H2O) and **AI Capabilities** (AGI, ASI, AN/WI).
- Expertise in **AI Functionalities** SLTR (S-Self Awareness, L- Limited Memory, T- Theory of Mind, R-Reactive Machine).
- Excellent in **Machine Learning Algorithms** (Linear & Logistic Regression, Decision Tree, Random Forest, SVM (Support Vector Machine), Naïve Bayes, K-means, KNN (K-Nearest Neighbors), Principal Component Analysis, Formal Detection, Dimensionally Reduction, Independent Component Analysis).
- Strong in **Machine Learning Types** (Supervised Learning (Classification, Regression), Unsupervised L (Clustering, Association), Semi Supervised L (Labeled Data, Unlabeled Data), Reinforcement L (Markov Decision Process).
- Expert in **Deep Learning Neural Networks, Neurons, Perceptron, BERT, NLP.**
- Good in **Natural Language Processing**: Tokenization, Stemming, Lemmatization, Corpus, Corpora.
- Expertise in **Data Visualization Tools**: **Tableau** and **Power BI Dashboard** Including create a various Charts, Plots, Map, Story and Graphs.
- Strong Hands-on in **ETL** Tools: **Informatica PowerCenter Transformations** (Active, Passive, Connected, Unconnected) and Talend Open- Studio.
- Good in **Robotics Process Automation Tools** (UiPath, Automation Anywhere, Blue Prism, Key Sight Egg Plant, PEGA)
- Strong in **Oracle SQL Database Schema Objects** (Table, View, Sequence, Keys plus Constraints, Synonyms, Materialized View, Operators, Indexes), Global Temporary Table & Plan_Table, Query Optimization.
- Expert in **SQL Statements**, Character, Date, Conversion, Aggregate & Group Functions, Conditions, Inner, Self, Cross & Outer Joins, Nested, Correlated & Inline View Sub-Queries, Pseudocolumns, DML MERGE, Materialized View (Refresh, Scheduling).
- Well Versed in **SQL Set Operators**, DML SIUD Commands & Statements, SQL*Plus Substitution Variable (&), Describe Command, Set Linesize, Set Pagesize.
- Excellent in **Oracle PLSQL Sub Programs** (Stored Function, Stored Procedure), Triggers, Cursors, Attributes, Packages (Spec, Body), Tuning Apps.
- Expert in **PLSQL Block Structures** (Named, Anonymous), **Pragma's**, Naming Conventions & Notations, **Error Reporting Functions** (SQLCODE, SQLERRM).

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Expertise in **Bill Inmon** and **Ralph Kimball Methodology** in **Data Warehouse Life Cycle**, **DWH Tables** (Dimension, Fact, Lookup, Fact less Fact Table), DWH Types Fact (Additive, Semi-Additive, Non-Additive), DWH Dimensions (Junk, Conformed, SCD).
- Good in 12 **E.F.Codd Rules** (NV-FISH-LID-PGDC), **Data Warehouse Dimensional Modelling** (Star Schema, Snow Flake Schema, Galaxy Schema), **ERwin Data Modelling** (CDM, LDM, PDM), **Entity Relationship** (1 to1, 1 to Many, Many to 1, Many to Many), **Normalization Techniques**.
- Excellent in **Big Data Engineer Spark, Scala, RDD** (Resilient Distributer Dataset), Hadoop, Sqoop, Mahout, Big Data (Diagnostic, Predictive, Descriptive, Prescriptive Analytics), and Big Data V's Volume, Value, Variety, Veracity, Velocity.
- Beginners in **Non-RDBMS:** Mongo DB, Julia DB, Maria DB, Dynamo & Couch DB.
- Strong in **SDLC Phases** and **SDLC Models** (**Waterfall, Agile Scrum**, V, Lean, Spiral, Rapid Application Development, Extreme Programming, Kanban.
- Knowledge in HTML, HTML5, CSS, CSS3, PHP, Bootstrap Framework, XML, Java Script used as Website Creation and Development.
- Good Knowledge in Data Structure & Algorithm, C, C++, VB Script, Visual Basic with MS Access, MS SQL Server, Oracle DB, Forms, Menus, Controls, Data source Name/DSN, ODBC, JDBC, Front end, Back end.

Technical Skills:

Data Science	Data Science Life Cycle, AI Tools and Frame works, AI Capabilities, AI Functionalities, ML Types, ML Algorithms, ML Life Cycle, Deep Learning, Neural Networks, Neurons, Perceptron, BERT, NLP
ML Algorithms	Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM (Support Vector Machine), Naïve Bayes, K-means, KNN (K-Nearest Neighbor), Principal Component Analysis, Formaly Detection, Dimensionally Reduction, Independent Component Analysis, Gradient Descent, Boosting (CAT, ADA, lightGBM, GBM)
Frameworks	Bootstrap, Django, Flask, Entity, MVC, MVP, MVVM
Project Management	SDLC Phases, SDLC Models, STLC Strategies, STLC Phases, STLC Techniques, UML Diagrams.
Software Languages	Oracle SQL, PLSQL Enterprise Edition 21c/19c, Python 3.11.0, R, SCALA, C, C++, VB
Technology	Data Science, AI, Big Data, Cloud Computing, Block Chain, Digital Marketing, Social Media Entertainment.
ETL Tools	Informatica Power Center 9.6.1, Talend Open Studio 8.0, Matillion
BI Reports	SSRS, IBM Cognos, Microsoft Strategy, Power BI Reports
Data Visualization Tools	Tableau 2022.4, MS Power BI
Markup Languages	HTML, HTML5, DHTML, CSS, CSS3
RPA Tools	AutomationAnywhere3.11.0,UiPath 2022.10.4,BluePrism 11.3
Operating System	MS DOS, MS Windows, HP Unix, Red hat Linux
MS Office Tools	Word, Excel, Powerpoint, Visio

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

Professional Projects Experience:
XecuteSmart Technologies, Data Science Lab, Theni / Chennai / Phoenix
Data Science Projects:
Environment: PYTHON 3.11.0, Anaconda JUPYTER Notebook 4.8, Modules/Libraries:
Pandas 1.4.2, NumPy 1.21.5, SK Learn 0.24, SciPy 1.7.3, Matplotlib 3.5.3, Seaborn 9.1, AI
Tools: TensorFlow 2.11.0, CNTK, PyTorch 1.13.0, Keras 2.11.0, ML Algorithms, DL, NLP

Life Science (Agri-Bio-Bot-Zoo)   **Oct 2021-Till Date**
Data Scientist
The Life Science comprise fields of science involving the study of living organisms such as plants, animals and humans. While biology remains the centerpiece of the life sciences, Technological advances in molecular biology and biotechnology have led to a burgeoning of specializations and new interdisciplinary fields. There are many disciplines included in the life sciences, however, such as anthropology, ecology, entomology, botany, zoology, microbiology, physiology, biotechnology, evolutionary biology, genetics, human anatomy, marine biology, molecular and cell biology, neuroscience, and biochemistry.
Responsibilities:

- Gathered Project requirements from Project Manager and Created The Research in **Life Science Project** from various domain such as **Botany, Zoology, Microbiology, Physiology, Biotechnology** in Evaluation and understanding the Project Requirements.
- Analyzed for **AI Tools Tensor Flow Framework** from **Speech Recognition**, **Expert System, Image Recognition** and **Machine Vision** in different application from project in discipline and characterized to animal and plant.
- Utilized **Machine Learning Algorithms** finding the animal and plant character some event **Supervised Learning** in **Classification** and **Regression** for correcting the problem **Un-Supervised Learning** feedback solving the problem.
- Developed by predicted the cells and **Biological** and **Psychology** in models scientifically implemented for **Python Packages** in library Pandas, NumPy, Seaborn, and Scikit learn, Plant and animals in cultivating the Unstructured data from life science project in customer linear Regression, Random Forest, SVM, Hierarchical Clustering.
- Improved the **Life Science** Project predict consumed problems in methods for solution in python implementation packages for finding the accuracy Matplotlip, NumPy using in Decision Tree, **Gradian Boosting Machine Learning algorithms,**
- Used **Vaccination** for humans. Bio Science project can be **AI Functionalities** Self Awareness, Limited Memory, Theory of mind, Reactive Machine in computational data for Solving the understanding Artificial intelligence benefits in cultivating seeds, human immunity power increasing, animals or birds behaviors.
- Built various **Machine Learning Techniques** and AI Tools & Frameworks for Dimensionality reduction, K-Means, Logistic Regression and Tensor flow, Pytorch in by **Increasing Productivity, Employment & Income** for all Domains in Life Science.

Healthcare + Insurance + Pharmacy + Medical Devices   **Dec 2021-Jan2023**
Data Scientist

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

The Healthcare Industry is an aggregation and integration of sectors within the economic system that provides goods and services to treat patients with curative, preventive, rehabilitative, and palliative care. It includes the generation and commercialization of goods and services lending themselves to maintaining and re-establishing health. The Healthcare Industry is comprised of companies that offer clinical services, manufacture drugs and medical equipment, and provide healthcare-related support services as Medical Insurance, It's also referred as Medical Industry.

Responsibilities:

- Collected **Healthcare Project** requirements from Project Manager. Design the Project various domain like Insurance, Pharmacy, Medical Devices in good understanding for requirements of Healthcare project.
- Utilized **Machine Learning Algorithm,** it can be Insurance Policy or Insurance Holder anything activity some Accident or Medical benefits applying insurance data cost evaluation and time periods of insurance company providing.
- Adapted **Machine Learning Algorithm** used in different Python Packages for pharmacy for drug in predicted diseases in patient take the order and correctly Medicine on the data in developed for **Python Libraries Pandas, NumPy, Scikit, Matplotlib.**
- Developed for identify the cancer patient sequential hospitalization like predicted correctly drug used in linear regression it makes Treatments carefully to patient.
- Improved the **AI Tools & Frameworks** Tensor Flow, Pytorch, CNTK, and Keras in ANN model for using PET scan, Eco in heart image detector in brain & heart disease identification. Used **Machine Learning Technique** for pharmacy Department went disease predicted what drug usage limit and measuring the values in control in disease performance data accumulated Applied in Machine learning algorithm **K- Nearest Neighbor, Decision Tree Classifier, Apriori, Naïve Bayes Algorithm.**
- Built in Medical Environment of Hospitalization it consumed by Healthcare in Pharm and Med device in improved by **Artificial Intelligence** in related with facilities working by various in RNN related with problems solved.
- Insurance for Customer Data it very clearly Machine Learning Technique used for **Supervised Learning** can be categorized in **Classification** and **Regression** problems. **Medical devices** Supervised learning can be used for those cases where we know the input as well as corresponding outputs.

Social Media Entertainment (SME)   **Apr 2022-Dec 2022**
Data Scientist

SME Projects refers to the means of interactions among people in which they create, share, and/or exchange information and ideas in virtual communities and networks. The Office of Communications and Marketing manages mainly The Facebook, Twitter, Instagram, LinkedIn, Whatsapp, and YouTube accounts. The use of entertainment media is often assumed to be motivated by individuals' desire to experience emotions. Entertainment audiences want experiences that make them laugh or cry, or keep them at the edge of their seats.

Responsibilities:

- Congregated **Social Media Entertainment** Project requirements from project manager. Making a project various domain like digital marketing, Facebook, Twitter, Instagram, Whatsapp Good and valuable information for SME project.

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Applied **DL Algorithms** Neural Network in Faster and easier to interpret large amount of data and meaning full information. Digital marketing in large amount of data various contribution if data communication and consumer for customer satisfaction of product.
- Used **ML Algorithm Technique** Decision Tree is a **Supervised Learning** Technique that can be used for both classification and Regression problems, It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions for variety situations in media.
- Adopted **ML Algorithm** in used **Principal Component Analysis** but mostly it is preferred for solving correlation, orthogonal, Eigenvectors, **Co-Variance Matrix problems.** Optimizing power allocation in various communication channels and image processing.
- Developed the various type of posters & promotion add for media entertainment sources prediction of stock market trends used in **DL Algorithms** for **Long Short Memory Networks (LSTMs).**
- Built in to streaming media, virtual reality gaming, and new delivery channels for news, music, and advertising. Developing for machine leaning algorithm supervised learning, Decision Tree, Classification Clustering Algorithm, Regression Tree Algorithm.
- Improved the **Artificial Intelligence Tool for NLP**, **CNTK** (Cognitive Tool Kit) it make to use for elaborate human brain setting to batch wise normalization and Sequence to Sequence Attention identify **CNN, RNN, FNN** component in **CNTK** and **NLP** to creating sequential process for order conformation marketing for **POS (Part of Speech).**
- **Social Media** created and sharing the Information, Marketing, Entertainment growth to economy for societies AI form of different applications Machine translation, Fighting Spam, Question-answering, **Sentiment Analysis.**
- Implemented to **Digital Marketing** for Digitalized Computer based Technology for sharing ideas and elaborate the communities using sentiment analysis, an application of **NLP** to identify the opinion and sentiment of their customers online.

AIRPA (Artificial Intelligence Robotics Process Automation)   June 2022 – Dec 2022
Data Scientist
The word "Robot" in Robotic Process Automation doesn't refer to a physical robot or an AI robot. It's a software robot, or bot, that automates mundane, repetitive tasks and processes, eliminating human error and increasing productivity and efficiency to a significant degree. RPA bots automate all rule-based tasks. Robotic Process Automation (RPA) is a Software Technology that makes it easy to build, deploy, and manage software robots that emulate humans actions interacting with digital systems and software.
Responsibilities:
- Congregated **AIRPA Project requirements** from Project Manager. Making a project various Tools and Direction, Process and working Good and valuable information**.**
- Applied the **BERT** (Bidirectional Encoders Replacement from Transformers) of Transformer in contrast to previous efforts which looked at a text sequence either from left to right or combined left-to-right and right-to-left training.
- Used to **BERT Sequential** operating for **Neural Network Model** containing includes robotics Automation process integrated with codes in **BERT Training** stages **MLM (Masked Language Model), NSP (Next Sentence Prediction).**

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Developed the **AI** provides robots with adequate computer vision and motion control to better understand the environment and act accordingly Rule Basis System, Context-awareness and Retention, Domain Specific Aptitude.
- **UiPath Software Robots** are intelligent, reliable, flexible, and eager to take on a range of tedious tasks. They're also easy to build and manage, so you can launch an entire robot workforce to quickly deliver efficiencies, higher performance, and high ROI.
- **Automation Anywhere** provides access control and compliance and that enables you to make sure that whoever is developing bots in **Automation 360** is following the organization's best practices, is storing their bots in a centralized place, and that enterprise-class encryption is being used to protect your bots/data both.
- **Blue Prism** is an **RPA Tool** which holds the capability of virtual workforce powered by software robots. This helps the enterprises to automate the business operations in an agile and cost-effective manner.
- **Built in Machine Learning Algorithm** Generative adversarial networks (GANs), Support vector machine, Dimensionality Reduction Technique in used to reconstruct 3D models of objects from images and model motion patterns in the video and data reduced, Signal processing dataset help in visualizing the data quickly.
- **Robotics Processing** Functioning software technology that makes it easy to build, deploy, and manage software robots that emulate humans actions interacting with digital systems and software **Deep Learning Algorithms in Neural Network CNN, LSTMs, SOMs.**
- Data Entry, Validation and Verification Robots for complete programing in **NLP Process** to Machine Translation, Inception, Information Retrieval, semantic analysis.

Environment: Robotics Process Automation Tools: Automation Anywhere 11.3, UiPath 2022.10.4, Blueprism 6.9, Orchestration

Blockchain Technology   **Jul 2022 – Dec 2022**
Data Scientist
Blockchain is a shared, immutable ledger that facilitates the process of recording transactions and tracking assets in a business network. Blockchain technology is an advanced database mechanism that allows transparent information sharing within a business network. Bitcoin is a cryptocurrency, while blockchain is a distributed database. Bitcoin is powered by blockchain technology, but blockchain has found many uses beyond Bitcoin. Bitcoin promotes anonymity, while blockchain is about transparency.

Responsibilities:
- Received various modules in project requirements and specifications from the project manager for **Blockchain Technology** Project Evaluate and Effective understanding the projects including **Bitcoin, Cryptocurrency** and **DLT (Distributed Ledger Technology).**
- Adapted **AI Tools** used **Pytorch** Dataset Type **MNIST** data base is image detected database in secure easy way to transforming Transfer to the programmable data immutable system for business network.
- Applied **BERT** is transferring containing the data participants agree to the validity of each of the records a language representation model that aims at tackling various **NLP** Tasks such as Question Answering, Language Inference, and Text Summarization.
- Used **Machine Learning Algorithm Technique** K-means is a Unsupervised Learning Technique that can be used for both clustering and Association problems, It is a graphical

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

representation for getting all the possible solutions to a problem/decision based on given conditions for various **Python Libraries** include Pandas, Numby, SciPy, Matplotlib.

- Built in **AI Tools** for **CNTK (Cognitive Tool Kit)** used dataset **IRIS flowers** properties Length, Width, Sepal, Petal in Blockchain technology part in DLT Distributed Ledger Technology system various steps include programmable, Secure, Anonymous, Unanimous, time stamped immutable, Distributed all steps utilize the **IRIS Dataset.**
- Developed **Machine Learning Algorithm** used for item sets in the transactional database and Select the minimum support to confidence in **Apriori Algorithm**, **Adversarial ML, Gradient Descent Algorithm.**
- Transaction for Currency, Information, Technique data and reduces the accuracy and performance of the model in **NLP Algorithms,** Machine Translation, Information Retrieval, **Word Level Analysis, Semantic Analysis.**
- **Retail Blockchain Technology** to anonymous, protecting the identities of the users. This makes Blockchain a more secure way to carry out Automated Transactions. Applied for **Deep Learning Algorithms Self Organizing Maps (SOM)**, Multilayer Perceptron's (MLPs), Radial Basis Function Network (RBFN).

Environment: Blockchain Technology, PYTHON 3.11.0, Anaconda Jupyter Notebook 4.8, Libraries: Pandas 1.4.2, Matplotlib 3.5.3, Seaborn 9.1, NumPy 1.21.5, SciPy 1.7.3, NLP, AI Tools: TensorFlow 2.11.0, CNTK, IRIS, Keras 2.7, PyTorch 1.13.0, ML Algorithms, NLP

STATAPPS (Statistical Applications)   **Aug 2022 – Jan 2023**
Data Scientist

Applied statistics is the root of data analysis, and the practice of applied statistics involves analyzing data to help define and determine organizational needs. Today we can find applied statistics in various fields. It is applied in marketing, e-commerce, banking, finance, human resource, production, and information technology. In addition, this mathematical discipline has been a prominent part of research and is widely used in data mining, medicine, A aerospace, robotics, psychology, and machine learning.

Responsibilities:

- Conglomeration various modules in **STATAPPS** application, uses of **AI, ML**, **Deep Learning, NLP** in project requirements and specifications from the project manager for Statistical Application project Evaluate and Effective understanding the projects.
- **Statistical Analysis** is to determine the cause-and-effect relationship in the analyzed data. Statistics works only on quantitative data and never on qualitative data.
- **Mean** (Average), **Median** (value repeating maximum number of times), **Mode** (central value of the total number of observations), **Standard Deviation (σ)** and **Variance (σ2)** in **Statistics** used to compare data. **Statistics** deals with every aspect of data, including the planning of data collection in terms of the design
- **Statistics** is at the core of sophisticated **Machine Learning** Algorithms, capturing and translating data patterns into actionable evidence in Linear Regression, Logistic Regression, Random Forest.
- **Statistical NLP** comprises all quantitative approaches to Automated Language Processing, including Probabilistic Modeling, Information Theory, and Linear Algebra **NLP** models Information retrieval, Question-Answering, Discourse processing.

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Application for identifying, collecting, organizing, summarizing, analyzing, interpreting, and presenting data for **AI Tools in Tensorflow** Mathematical concepts Vector Space Model, Mathematical Computation.
- Developed in transforming the **Data Science Deep Learning algorithms** for **Multilayer Perceptron (MLPs), Self-Organization Maps (SOMs), Recurrent Neural Network (RNNs)** Statistics is a crucial process behind how we make discoveries in science, make decisions based on data, and make predictions.
- **Statistical Application** Mathematical Science for analyzing existing data pertaining to specific problems, applying statistical tools to this data, and presenting the results for decision-making. Evaluated the predicted accuracy of several **ML** approaches prior to selecting the most accurate simple model as **Linear Regression** is generally the starting point for a **Statistical Analysis** & consistency of data is evaluated to determine if satisfies the model hypothesis.

Environment: R Programming, R Studio 4.1.3, Arithmetic Mean, Median, Mode, SD/Standard Deviation, Variance, ANOVA, ANCOVA, IRIS Flower Dataset, Data Visualization, SAS, ML

Data Analyst Projects:
Environment: Oracle 21c enterprise edition, OLTP, OLAP, Informatica power center 10.5, Talend 8.0.1, SSAS, SSRS, Tableau 2022.4, MS Power BI, Dr. E.F. Codd 12 rules.

BankFinMort (Banking + Finance + Mortgage)   **Dec 2021 – Jun 2022**
Data Analyst
Bank is a Financial Institution that is licensed to accept checking and savings deposits and make loans. Banks also provide related services such as Individual Retirement Accounts (IRAs), Certificates of Deposit (CDs), currency exchange, and safe deposit boxes. Finance, of financing, is the process of raising funds or capital for any kind of expenditure. It is the process of channeling various funds in the form of credit, loans, or invested capital to those economic entities that most need them or can put them to the most productive use. A Mortgage is a loan from a bank or other financial institution that helps a borrower purchase a home. The collateral for the mortgage is the home itself. That means if the borrower doesn't make monthly payments to the lender and defaults on the loan, the lender can sell the home and recoup its money.

Responsibilities:
- Congregated the **BankFinMort Projects** requirements from the project manager and performed strong analysis to understand the subjects.
- Composed different types of **Logical, Physical Data Modelling** to understand the entity relationship as follows (1 to 1, 1 to Many, Many to 1, Many to Many) between **100+ tables**
- Planned **Dimension Modelling** like **Star, Snowflake, Galaxy** based on the requirements.
- Enforced **Dr.E.F.Codd 12 Rules** such as **Logical Data** Independence rule, Information rule, Physical data Independence rule for the project preparation based on requirements
- Executed various **SQL Queries/Statements: (SIUD-Select, Insert, Update, Delete), (CAD-Create, Alter, Drop/TRC- Truncate, Rename, Comment)** and **SQL Joins** such as Inner, Outer, Full outer, Right outer, Cross Joins utilized **Materialized View** based on the Project requirements.
- Improved **PL/ SQL Programming** which includes Stored procedure, Stored Function, **PL/SQL Packages** (Package Specification, Package Body), **Triggers**, **Cursors** for **BankFinMort Modules.**

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Utilized **ETL Informatica Powercenter Tool** and Adopted necessary transformation (Active/Connected, Passive/Connected, Active/Unconnected, Passive/Unconnected) like Aggregator, Expression, Filter, Joiner Transformations, Rank Transformations and **Talend ETL Tool Transformations** tFilter Row, tDenormalize, tLoop, tJoin, tNomalize, tReplicate, tUnique Row.
- Explored **SSIS** (SQL Server Integration Service) and **SSRS** (SQL Server Reporting Service) Such as PASS-LCD-CD.
- Designed and Developed **Tableau Tool** in **Data Visualization** for visually analyzing the data. Drawn various **Plots** and **Charts** (Bubble, Bump, Bar, Donut, Pie, Box Plot, Dual Lines, Pareto Plot, Circle Views) to Visualize Based on Requirements.

Smart Reservation System   **Oct 2021 – Apr 2022**
Data Analyst
SRS Projects are computerized system used to store and retrieve information, conducting transactions related to bus, train, air travel, hotels, food, car rental, or other activities. Reservation Management System is a cloud-based tool to help you manage your property's operations and bookings to securely collect and store your details. Serving as a centralized management system, a Reservation Management System allows you to keep track of everything in one place.
Responsibilities:

- Gathered **Smart Reservation System** in Different Requirement and specification from the Project Manager to develop the Various types of Reservation Projects.
- Executed Analysis to **CDM, LDM, PDM** in ERwin Tool for ER Diagram relationship Such as one to one, One to Many, Many to One, Many to Many.
- Improved **Data Modeling, Dimension Modeling** utilized for Star schema, Snowflake schema, Galaxy schema knowledge based on requirement and collecting the checklist definitions for **150 Plus Tables** on dimension modeling.
- Practiced by Different **SRS Modules** for **Dr.E.F.Codd 12 Rules** contain Systematic Treatment of NULL value, Guaranteed Access, Comprehensive Data Sub-Language, Distribution Independence.
- Manageable Multiple **SQL Keys, Constraints**, sequence, operators (set operators, logical operators, relational operators), single row functions (number, date, conversion, transformation, Miscellaneous Function), Pseudo columns.
- Applied for **PL/SQL Programs** on Anonymous and Named Pl/SQL Block Structure, Exceptions, Control Structures, Sub-Programs, Stored Function, Stored Procedure.
- Enforced Various type of **ETL Transformations** by using **Informatica Power Center** Tool like Sorter, Lookup, Rank, Router, Source qualifier and tNormalizer, tFilelist, tFilter Row, tAggregate row, tJavaRow, tJoin, tunite in tranformations in **ETL Talend Tool.**
- Explored **SSRS Reports** such as Parameterized, Adhoc, Sub, Snapshot, Linked, Cache, Drill through, Click Through, Drill Down.
- Designed by Analyzed in **Data Visualization** Tools: **Power BI & Tableau Tool**. Built Different Types of Charts and Plots using Funnel, Area, Gantt, Box Plot, Bullet Graph, Lolli Pop Chart, Scatter Plot.

Human Resource Information System   **Nov 2021 – Jun 2022**
Data Analyst

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

Human Resource Information System (HRIS) is a software package to aid human resources professionals in managing data. HR Professionals Utilize these systems to facilitate work flow, Improve Efficiency, Store and Collect Information. Several companies offer HRIS packages to Employers. HRIS applications includes an Employee Self-service portal, Payroll, Workforce Management, Recruitment and Hiring, Benefits Administration, and Talent Management.

Responsibilities:

- Collected Various **Project Modules Requirements** and **Specification** from the Project Manager **HRIS Package Project.**
- Utilized Data Modeling Physical, Logical and Conceptual data modeling diagram like Erwin data modeling tool Using entity Relationship various type (1 to 1, 1 to many, many to many, many to 1) between **100+ tables**
- Planned **Dimensional Modeling** such as **Star Schema, Snowflake schema, Galaxy Schema** in using all collect the based-on project Requirements
- Evaluated the all-checklist definitions in **Data Modeling & Dimensional Modeling** based for Project knowledge
- Used **Dr.E.F.Codd rules** like Non Subversion Rule, High Level Insert, Update, Delete Rule, Distribution independence Rule, Comprehensive Data sub-language rule.
- Manipulated Different **SQL Simple & Complex Queries** as follows **SQL Subqueries** utilized **Simple, Inline view, Scalar, Nested & Correlated Subquery**, used different **Data Types, Tables** (Dual, Plan, Global Temporary Table), **DML MERGE Command**.
- Developed **PL/SQL Programming** which contains **Stored Procedure, Stored Function,** Records, Triggers, **Cursors** and various **Types of Pragma's**
- Applied different **Types of Transformations** for **ETL Informatica Power center** Tool used for SQL Transformation, Union Transformation, XML Generator, Update Strategy Transformations and Talend open Studio transformations Such as tDB Row, tConvert Type, tBuffer Input, tFixed Input, tBuffer Output,tMap Joint.
- Built **Data Visualization** to Analyze the data using Power BI and Tableau. Created different types of Charts as Plots Bar Chart, Pie, Donut, Bump, Box Plot, Whisker Plot, Pareto Plot in based on the project requirements.
- Explored SQL Server Reporting Services Such as Parametrized report, Adhoc Report, Sub Report, Snapshot Report, linked Report, Cache Report, Drill Through, Click Through, Drill Down for based on project reporting services.

Supply Chain Integration   **Oct 2021 – Mar 2022**
Data Analyst

Supply Chain Integration essentially means that the information and communication systems of all stakeholders are able to seamlessly exchange information through all planning, execution and completion of transport and logistics operations throughout a product's life time. Supply Chain Management is the practice of coordinating the various activities necessary to produce and deliver goods and services to a business's customers.

Responsibilities:

- Received various modules in **Project Requirements** and **Specifications** from the Project Manager for **Supply Chain Integration Project,** Evaluated and Effective Analysis to understand the subjects.

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Used **ERwin Tool** to planned **Conceptual** Data Modelling, **Physical** Data Modelling, **Logical** Data Modelling to understand the **ER Diagram** as follows (one to one, one to many, many to many, many to one) relationship between **100 plus Tables**
- Designed **Dimensional Modelling & Data Modelling** for based Knowledge for project check list definitions in DWH Dimensional Modelling **Star, Snowflake, Galaxy.**
- Utilized **Dr.E.F.Codd Rules** as follows Supply Chain Integration Project modules Guaranteed Access, Information, Integrity Updating, Comprehensive Data Sub-language.
- Developed Various **SQL Simple & Complex Queries** like Create synonyms for tables, SQL Expressions follows Simple, Compound, Arithmetic, Boolean and constrains +keys, Table level Constrains, Column Level Constrains.
- Established **PL/SQL Programming: Cursor** (Implicit, Explicit), **Control Structures** (Condition, Iterative, Sequential Controls), **Package** (Package Specification, Package Body), **Sub-Programs** as follows Stored Procedures and Stored Functions.
- Executed **ETL Talend Tool** different **Transformations** tJava Row, tFilter Columns, tLoop, tFuzzymatch, tReplace, tRun Job, tSample Row and Expression Transformation, Sorter Transformation, Rank Transformation, XML Source Qualifier Transformation, Lookup Transformation, Java Transformation in ETL Informatica Power Center.
- Adopted **SSRS Reports** such as PASS-LCD-CD (P-Parametrized Report, A-Adhoc Report, S-Sub Report, S-Snapshot Report, L-Linked Report, C-Cache Report, D-Drill Through, C-Click Through, D-Drill Down) based on Project Requirement.
- Explored **Data Visualization** to Analyze Data Using **PowerBI** and **Tableau Tool**. Create various Plots and Charts: Dual Lines chart, Pie Chart, Bubble Chart, Donut Chart, Water Fall Chart in Tableau Tool and visualize the dash board and reporting in **PowerBI Tool** visualize the data based on customer requirements.

BIGDATA Projects:

Environment: BigData, Hadoop 3.3.4, HDFS Arch, SQOOP, SEMMA, CRISP –DM, Flume, Spark 3.3.1, Scala 3.1.1, Hive 3.1.2, Kafka 3.1.0, Data Bricks 11.0, PySpark 3.1.1.

FOREX (Foreign Exchange)   **Aug 2022 – Jan 2023**
Data Engineer
The Foreign Exchange Market is a Global decentralized or over-the-counter market for the trading of currencies. This market determines foreign exchange rates for every currency. It includes all aspects of buying, selling and exchanging currencies at current or determined prices. These buyers use the foreign exchange to pay for imports of merchandise, to make service payments (inclusive of travel), to repatriate profits, and to repay external debt.

Responsibilities:

- Gathered **Forex (Foreign Exchange**) project requirement from project manager. Evaluate the understanding project approaches in application predicted foreign exchange project.
- Applied for **Bigdata V's** in volume, velocity, variety, veracity, value in project data size of volume, speed of change, form of data source, Uncertainly data, collect value of data from business. Improved Transaction Process System **Hadoop Efficiency** Storage of Dataset process to multiple computer systems in clustering. Developed the exchange rate, the price of one currency in terms of another, helps to determine a nation's economic health and hence the well-being of all the people residing in it.

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Implemented **MapReduce** for filtering, sorting data sets parallel distributed algorithms and Sqoop, spark command line interface application used for big data workloads memory catching and optimized Query execution.
- **Dataframes, Datasets & SQL** provide the higher-level abstraction over **RDD** (Resilient Distributed Dataset) foreign exchange market is a global decentralized or over-the-counter market for the trading of currencies.
- **CRISP-DM Methodology** used to tackle problems in traditional BI Data Mining and **SEMMA** is another **Methodology** developed by **SAS** for data mining modeling it's used for **Bigdata Environment.**
- Built in **SCALA**, Databricks both **Object-Oriented** and Functional **Programming** used in **Data Processing**, distributed computing, web development, data analysis with a simple collaborative environment to run interactive and scheduled data analysis workloads.
- Because of the worldwide reach of trade, commerce, and finance, forex markets tend to be the largest and most liquid asset markets in the world. Currencies trade against each other as exchange rate pairs.

Clinical Trials   **Sep 2022 – Feb 2023**
Data Engineer
Clinical trials are research studies performed in people that are aimed at evaluating a medical, surgical, or behavioral intervention. They are the primary way that researchers find out if a new treatment, like a new drug or diet or medical device is safe and effective in people. Often a clinical trial is used to learn if a new treatment is more effective and/or has less harmful side effects than the standard treatment. Clinical trials are a type of research that studies new tests and treatments and evaluates their effects on human health outcomes.

Responsibilities:
- Collected understanding and Evaluated to various Application for project. **Clinical Trial** Project in **Medical, Surgical,** other operation Project requirement from Project Manager.
- Adapted in treatment of human and animal for primary new drug and medicine checkup in data analyzed, data mining, Modeling for Effective **SEMMA, CRISP Methodology** used.
- **Clinical Trials** are carefully designed, reviewed and completed, and need to be approved before they can start for some analytics conducted in Predictive, Descriptive, Cognitive, Diagnostic, Prescriptive Analytics.
- Developed **Bigdata** Key terms in **V's** volume- patient data size, velocity- drug or medicine speed or effective, variety – patient report data sources, veracity- data prediction of future events, value – data collated in detailed for patient.
- Applied for more storage in clinical Data set from **GFS (Google File System), HDFS (Hadoop Distributed File System)** it human for effective and continuous various methods for clinical test in this example for new Medicine for cancer patients
- Built in **Map Reduce, SQOOP** in Method Utilized to Big datasets with performs for Filtering, Sorting Data set **Parallel Distributed Algorithm** on a cluster and Transferring data from data base system. **Clinical Trials** require more time and attention than standard treatment would, including visits to the study site, more blood tests, more procedures, hospital stays, or complex dosage schedules in conducted data **RDD (Resilient Distributed Dataset)** immutable data set collected Key terms used an **RDD** Transforms, Actions Record data on multiple nodes

Er. G.Vignesh Karthik, BE, MBA
Email: gvigneshkarthik.datascientist@gmail.com
Contact: 89407 29161

- Implemented **Datasets** can be Evaluated **SPARK, SCALA** in code Utilizes memory Catching, optimized Query Execution for any size of data Researchers test possible new drugs in the laboratory to begin with.
- **Clinical Trials** is to establish the effect of an intervention. Treatment effects are efficiently isolated by controlling for bias and confounding and by minimizing variation. Screening and checking the participant whether he/she qualifies for the test.
- **Bigdata** in Clinical study refers to the information collected using electronic database. **Data Bricks** Data come from daily routine clinical practice without modification or screening with strict inclusion and exclusion criteria, therefore retaining its real-world.

Other Projects: **Oct 2021 – May 2022**

C, C++ Programming, Website Design

C is a General-purpose Programming Lang, It's very powerful; it has been used to develop OS, DB, Apps, etc. C – Comments, Variables, Data Types, Integer, Float, Character, Conversion, Constant, Operators, Break /Continue, While, For, Switch, Strings, Characters, Functions, Arrays, User Input

C++ - Programming: C++ is a popular programming language. It is used to create computer programs, and is the most used language in game development. OOP, Classes/Objects, Methods, Constructors, Access Specifiers, Encapsulation, Inheritance, Polymorphism, Files, Exceptions

Data Structure Algorithm, DSA - Algorithms Basics, Asymptotic Analysis, Greedy Algorithms, Divide and Conquer, Dynamic Programming, **Stack & Queue:** DSA – Stack, Expression Parsing, Queue, **Searching Techniques-**DSA – Search: Linear, Binary, Interpolation, Hash Table, **Sorting Techniques:** DSA - Sorting Algorithms, Sorts: Bubble, Insertion, Selection, Merge, Shell Quick

Website Creation: xecutesmarttechnologies.com - HTML, HTML5, CSS, CSS3, PHP, XML, JS, Bootstrap Framework - Determine the type of your Website, Determine a Domain name, Determine the Platform to use, Register Domain name, Sign up for Web hosting, Point your Domain name to your Web hosting, Build the Website

Educational Qualification and Training Credentials:

- Master of Business Administration (HR & Finance) -7.15
  RVS College of Engineering & Technology - **Anna University 2019**
- Bachelor of Engineering (Mechanical) – 6.55 –
  VPV College of Engineering – **Anna University 2015**
- Diploma in Mechanical Engineering – 70.27%
  Alagappa Polytechnic College – **DOTE 2012**
- Auto CAD, Pro-E., **CADD Centre, Dindigul - 2012**
- Worked as a **Technical Assistant,** Corrosion department Polymer preparation and technical work. Cutting and polishing the M.S. Sheet. Etc., **CECRI-CSIR, Karaikudi** - 5.8.2009 - 31.5.2011
- Apprentice Training Fabrication of soft are hard tolls for different experiments. CNC training, Lathe and Milling., **CECRI-CSIR, Karaikudi** - 11-04-2008 - 10-04-2009
- Worked as a **Technical Assistant,** Fabrication of reaction vessels, graphite crucibles, arrangements of Anodes and cathodes, High temperature Handing etc., **CECRI-CSIR, Karaikudi** 10.04.2007-10.04.2008.