INTERNSHIP REPORT

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

SUMEER SHARMA

19BCS6018

AT
HIGHRADIUS

Under the Supervision of :

MR. RITESH RAJ SINGH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

APEX INSTITUE OF TECHNOLOGY

CHANDIGARH UNIVERSITY, GHARUAN,
MOHALI - 140413, PUNJAB

MAY & 2022

A. Distribution of hours:

Orientation **___1 Hour____**

Observing **___1.5 Hour____**

Meetings (e.g. staffing, working with the team, etc) **___460 Hours____**

Lectures, Seminars, Conferences**___20 Hours____**

Assessment **___10 Hours____**

Planning (activity analysis, goals and objectives, etc) **__20 Hours___**

Studying/Researching **__460 Hours___**

B. Implementation (in *hours* which so ever is applicable. Otherwise mention Not Applicable):

a. Leadership **__NA___**

b. Counselling **__NA__**

c. Supervision **__NA__**

d. Evaluation **___NA___**

e. Documentation **___NA___**

f. Discharge/Transition Plans **___NA___**

g. Other (Please specify) **___NA___**

C. Total clock hours during this report period : __600 Hours__

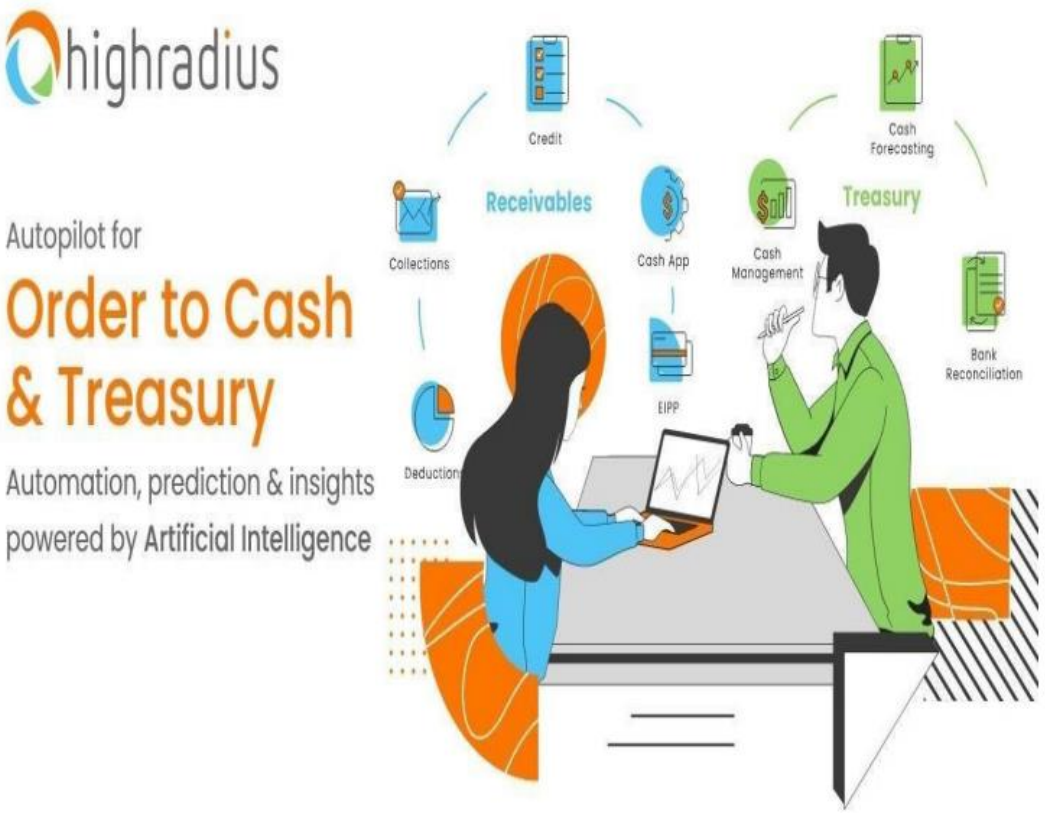
INDEX :

INTRODUCTION TO COMPANY AND INDUSTRY :

The journey to the High radius started in our university placement cell wherever we found the links we wanted and were able to register and then the company an orientation webinar where all the details were given to us like how things will go and what will be the life we we got selected at HighRadius. The journey began with an accurate check that encapsulated some of the circular writing skills and basic communication skills. Due to the current covid-crisis situation like all MNC we are given a working environment at home yet by working with mentors and team leaders we are often able to achieve current goals with our training value. they are given all the required documents and all the main categories start on some days. HighRadius offers cloud-based Autonomous software based in the CFO workplace. The world's top 700 firms have reorganized their order in finance, treasury and records to report processes through HighRadius. Our customers include 3M, Unilever, Anheuser-Busch InBev, Sanofi, Kellogg Company, Danone, Hershey's and many more.

Autonomous software system is a data-driven software program that constantly changes its behavior to a sub-domain data transformation. It introduces common digital transformation capabilities such as Artificial Intelligence, Robotic Automation, the process of language communication Participate in Connected Workplaces as out-of-the-box financial and accounting options. Participants in the financial business believe that they need only 2 options: decide on an Excel based

software vendor or process based on an electronic record system, or, choose an AI or RPA middleware platform to create and maintain. internally, domain-specific skills. In contrast, the HighRadius Autonomous software system combines the best of each world to deliver measurable business results that recommend DSO reduction, asset improvement, bad debt reduction, monthly closing time and product improvement in less than six months. Our sales offer value to a wide range of major customers | consumers} and are very important in industries such as consumer products, manufacturing, distribution, power, different selling products or providing service to other businesses. Our clients come from many of the world's most important companies and over 200+ firms worldwide 2000 moreover as medium-sized enterprises with no IT resources to integrate into the Associate in Nursing ERP platform yet still need to do their own and contour. their assets and wealth processes. Power and Productivity enhancements are essential to what HighRadius provides to our customers. regardless of which ERP system, assets or Treasury Management system you use, our sales automatically perform complex tasks, communicating and directing, and allow for process processes to run the best acquisition processes and treasury processes. we often empower our customers to be more efficient and effective, predict and manage cash, prompt payments, and improve key metrics like Days Sales Outstanding (DSO) and improve product availability.



The HighRadius culture must continue to work in completely different ways to help everyone grow. Let the most effective program win Crowdsource by using the 1st Principles method to make advanced choices A bachelor's degree in Boss Transparency from above builds the trust of everyone Be courageous and straightforward Response is part of a permanent business debate Gritty or home grit is the # 1 reason for fulfilment. Hope in the roller coaster Fail-speed, read-speed, quick-fix but they do Humble yourself but be able to stir up curiosity and personality that we often learn and grow professionally Ride or Die with a client, customer service is everyone's responsibility "Integrity - we work with customers and trust with openness, honesty and integrity" - was the key to a notorious, collapsed energy company. you may notice several firms out there where ideas and values are finalized in the jargon of a leading company and what you see is not what you get

DETAILED INTRODUCTION TO YOUR JOB PROFILE :

It is 3 Months Internship at Highradius Corporation in which I have been working on a project AI Enabled Invoice Management System. Due to Covid-19 pandemic in India this internship is virtual So I have to attend all the scrum calls and meetings virtually through Zoom app.

In this Internship we made a application to help the people working in the Accounts Receivable departments in their day-to-day activities. We build a web application where the users in the Account Receivable department can View the invoice data from various buyers, See various fields/attributes of the invoice(s) from a particular buyer , Perform Data Pre-processing on the invoice data , Get account-level analytics to easily visualize and interpret data - EDA and Feature Engineering , Get a prediction of when the invoice is going to get paid. The objective of the Web Application Development internship project is : To build a Full-stack Invoice Management Application using ReactJs, JDBC, Java, Servlets , Build a responsive Receivables Dashboard , Visualize Data in the form of grids, Visualize Data in the form of graphs. Perform Searching operations on the invoices. Add & Edit data in the editable fields of the grid. Delete data of selected rows in predefined templates.

In this Internship we have to build an AI Enabled Fintech B2B Invoice Management Application.

Introduction to B2B Operations :

The B2B world operates differently from the B2C or C2C world. Businesses

work with other businesses on credit. When a buyer business orders goods

from the seller business, the seller business issues an invoice for the same.

This invoice for the goods contains various information like the details of the goods purchased and when it should be paid. This is known in accounting terminology as “Accounts Receivable”.

“Accounts Receivable represents money owed by entities to the firm on the sale of products or services on credit. In most business entities, accounts receivable is typically executed by generating an invoice and either mailing or electronically delivering it to the customer, who, in turn, must pay it within an established timeframe, called credit terms or payment terms.”

Seller business interacts with various businesses and sells goods to all of them at various times. Hence, the seller business needs to keep track of the total amount it owes from all the buyers. This involves keeping track of all invoices from all the buyers. Each invoice will have various important fields like a payment due date, invoice date, invoice amount, baseline date etc.

The buyer business needs to clear its amount due before the due date. However, in real-world scenarios, the invoices are not always cleared ie. paid in full amount by the due date. The date on which a customer clears the payment for an invoice is called the payment date.

Account receivables Department :

1. In the ideal world, the buyer business should pay back within the stipulated time (i.e. the Payment Term). However, in the real world, the

buyer business seldom pays within their established time frame, and this is where the Account Receivables Department comes into the picture.

2. Every business consists of a dedicated Account receivables Department to collect and track payment of invoices.

3. It consists of an Account receivables team that is responsible for :

- Collecting payments from customers for their past due to invoices.
- Sending reminders and follow-ups to the customers for payments to be made.
- Looking after the entire process of getting the cash inflow.
- Help the company get paid for the services and products supplied.

About Machine Learning Part :

As a winter internship project, you will be building a web application to help the people working in the Accounts Receivable departments in their day-to-day activities. You need to build a web application where the users in the Account Receivable department can :

- View the invoice data from various buyers.
- See various fields/attributes of the invoice(s) from a particular buyer.
- Perform Data Pre-processing on the invoice data.
- Get account-level analytics to easily visualize and interpret data- EDA and Feature Engineering.
- Get a prediction of when the invoice is going to get paid.

About Web Application Development Part :

The objective of the Web Application Development internship project is :

- To build a Full-stack Invoice Management Application using ReactJs, JDBC, Java, Servlets.
- Build a responsive Receivables Dashboard.
- Visualize Data in the form of grids.
- Visualize Data in the form of graphs.
- Perform Searching operations on the invoices.
- Add & Edit data in the editable fields of the grid.
- Delete data of selected rows in predefined templates.

The mandatory features are compulsory tasks and the optional features are for extra credit points, which will give you an added advantage. The mandatory features for our webapp is UI Creation , Grid Creation , Crud Operations , Pagination and Advance Search.

Technical stacks used:

• Machine Learning (Python)

• Java

• React

• SQL Machine Learning

Machine Learning : In our first Milestone project we were given a data in which we have to predict the clear date for companies using Machine Learning Algorithms and ML modelling. Low-cost data-driven management decisions lead to a different management style, in which insurance leaders and future bank agents will ask the right questions on the machines, rather than on human experts. The machines will then analyze the data and come up with recommended results, which

can help leaders and subordinates make better decisions. Customers looking at systems such as text chats, voice systems or Chatbot Finance can provide a customized personal service or expert advice experience at a lower cost. Mathematical tools gather evidence and analyze the data needed to convict. Artificial Intelligence tools then learn and monitor user behavior patterns to detect unusual signs and warnings of fraudulent attempts and incidents. Claims management can be set up using ML Techniques (ML) at different stages of the claim process. By using Artificial Intelligence and managing large amounts of data in a short period of time, insurers can create their own management system. It can even bind certain claims, reduce all processing time and hosting costs while improving customer experience. These algorithms identify patterns in data to help identify fraudulent claims in the process. With their self-study skills, Artificial Intelligence (AI) systems can then adapt to new unfamiliar situations and improve discovery over time. It can serve as a game changer by improving business performance, improving internal processes, and surpassing competitors. Analytics works closely with organizations in many industries to collect and organize data, analyze it using our advanced algorithms and technologies and quickly deploy customizable, decision-making solutions that are unique to each client.

NumPy : It is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy. Arrays are very frequently used in data science, where speed and resources are very important.

Pandas : It is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another

package named Numpy, which provides support for multi-dimensional arrays. As one of the most popular data wrangling packages, Pandas works well with many other data science modules inside the Python ecosystem, and is typically included in every Python distribution, from those that come with your operating system to commercial vendor distributions like ActiveState's ActivePython. Pandas makes it simple to do many of the time consuming, repetitive tasks associated with working with data, including: Data cleansing Data fill Data normalization Merges and joins Data visualization Statistical analysis Data inspection Loading and saving data

Data preprocessing : Data preprocessing involves transforming raw data to well-formed data sets so that data mining analytics can be applied. Raw data is often incomplete and has inconsistent formatting. Data goes through a series of steps during preprocessing:

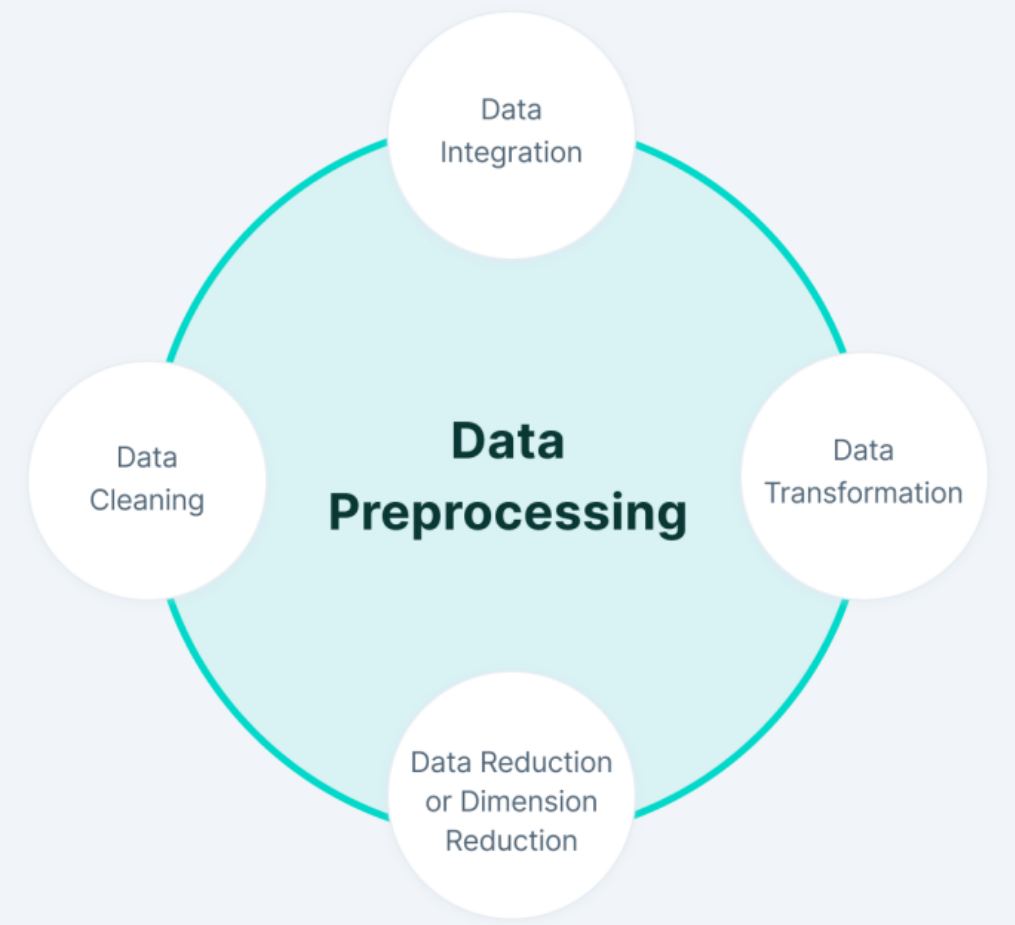
Data Cleaning : Data is cleansed through processes such as filling in missing values or deleting rows with missing data, smoothing the noisy data, or resolving the inconsistencies in the data. Smoothing noisy data is particularly important for ML datasets, since machines cannot make use of data they cannot interpret. Data can be cleaned by dividing it into equal size segments that are thus smoothed (binning), by fitting it to a linear or multiple regression function (regression), or by grouping it into clusters of similar data (clustering). Data inconsistencies can occur due to human errors (the information was stored in a wrong field). Duplicated values should be removed through deduplication to avoid giving that data object an advantage (bias).

Data Integration: Data with different representations are put together and conflicts within the data are resolved.

Data Transformation: Data is normalized and generalized. Normalization is a process that ensures that no data is redundant, it is all stored in a single place, and all the dependencies are logical.

Data Reduction: When the volume of data is huge, databases can become slower, costly to access, and challenging to properly store. Data reduction aims to present a reduced representation of the data in a data warehouse. There are various methods to reduce data. For example, once a subset of relevant attributes is chosen for its significance, anything below a given level is discarded. Encoding mechanisms can be used to reduce the size of data as well. If all original data can be recovered after compression, the operation is labeled as lossless. If some data is lost, then it's called a lossy reduction. Aggregation can also be used to condense countless transactions into a single weekly or monthly value, significantly reducing the number of data objects.

Data Discretization: Data could also be discretized to replace raw values with interval levels. This step involves the reduction of a number of values of a continuous attribute by dividing the range of attribute intervals. Data Sampling: Sometimes, due to time, storage or memory constraints, a dataset is too big or too complex to be worked with. Sampling techniques can be used to select and work with just a subset of the dataset, provided that it has approximately the same properties of the original one.



EDA

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better EDA tends to work with different libraries involved in ML as mentioned above EDA is also breaked into different parts such asReading data- we use a CSV file containing all the raw data of a certain company or retail transaction like how when and where all these were done. Descriptive Statistics- By assigning include attribute a value of ‘all’, we make sure that categorical features are also included in the result. Missing value imputation- This tell us how many missing values we have in each column in our dataset Graphical

representation- All the data is aligned into different types of graphs such as bar graph which makes it more easy to understand When building a machine learning model in real-life, it's almost rare that all the variables in the dataset are useful to build a model. Adding redundant variables reduces the generalization capability of the model and may also reduce the overall accuracy of a classifier. Furthermore adding more and more variables to a model increases the overall complexity of the model. Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set. It can produce new features for both supervised and unsupervised learning, with the goal of simplifying and speeding up data transformations while also enhancing model accuracy. Feature engineering is required when working with machine learning models. Regardless of the data or architecture, a terrible feature will have a direct impact on your model. Feature Creation: Creating features involves creating new variables which will be most helpful for our model. This can be adding or removing some features. As we saw above, the cost per sq. ft column was a feature creation. Transformations: Feature transformation is simply a function that transforms features from one representation to another. The goal here is to plot and visualise data, if something is not adding up with the new features we can reduce the number of features used, speed up training, or increase the accuracy of a certain model. Feature Extraction: Feature extraction is the process of extracting features from a data set to identify useful information. Without distorting the original relationships or significant information, this compresses the amount of data into manageable quantities for algorithms to process. Benchmark : A Benchmark Model is the most user-friendly, dependable, transparent, and interpretable model against which you can measure your own. It's a good idea to run test datasets to see if your new machine learning model outperforms a recognised benchmark. These benchmarks are often used as measures for comparing the performance between different machine learning models like neural networks and support vector machines, linear and non-linear classifiers, or different approaches like bagging and boosting. To learn more about feature engineering steps and process,

check the links provided at the end of this article. Now, let’s have a look at why we need feature engineering in machine learning. 1. Null Imputation When it comes to preparing your data for machine learning, missing values are one of the most typical issues. Human errors, data flow interruptions, privacy concerns, and other factors could all contribute to missing values. Missing values have an impact on the performance of machine learning models for whatever cause. The main goal of imputation is to handle these missing values. There are two types of imputation : Numerical Imputation: To figure out what numbers should be assigned to people currently in the population, we usually use data from completed surveys or censuses. These data sets can include information about how many people eat different types of food, whether they live in a city or country with a cold climate, and how much they earn every year. That is why numerical imputation is used to fill gaps in surveys or censuses when certain pieces of information are missing. #Filling all missing values with 0 data = data.fillna(0) Categorical Imputation: When dealing with categorical columns, replacing missing values with the highest value in the column is a smart solution. However, if you believe the values in the column are evenly distributed and there is no dominating value, imputing a category like “Other” would be a better choice, as your imputation is more likely to converge to a random selection in this scenario . #Max fill function for categorical columns

data[‘column_name’].fillna(data[‘column_name’].value_counts().idxmax(), inplace=True) 2. Handling Outliers Outlier handling is a technique for removing outliers from a dataset. This method can be used on a variety of scales to produce a more accurate data representation. This has an impact on the model’s performance. Depending on the model, the effect could be large or minimal; for example, linear regression is particularly susceptible to outliers. This procedure should be completed prior to model training. The various methods of handling outliers include: Removal: Outlier-containing entries are deleted from the distribution. However, if there are outliers across numerous variables, this strategy may result in a big chunk of the datasheet being missed. Replacing values: Alternatively, the

outliers could be handled as missing values and replaced with suitable imputation. Capping: Using an arbitrary value or a value from a variable distribution to replace the maximum and minimum values .

Discretization : Discretization is the process of converting continuous variables, models, and functions into discrete ones. This is accomplished by constructing a series of continuous intervals (or bins) that span the range of our desired variable/model/function. 3. Log Transform Log Transform is the most used technique among data scientists. It's mostly used to turn a skewed distribution into a normal or less-skewed distribution. We take the log of the values in a column and utilise those values as the column in this transform. It is used to handle confusing data, and the data becomes more approximative to normal applications. //Log Example df[log_price] = np.log(df['Price']) 4. One-hot encoding A one-hot encoding is a type of encoding in which an element of a finite set is represented by the index in that set, where only one element has its index set to "1" and all other elements are assigned indices within the range [0, n-1]. In contrast to binary encoding schemes, where each bit can represent 2 values (i.e. 0 and 1), this scheme assigns a unique value for each possible case. 5. Scaling Feature scaling is one of the most pervasive and difficult problems in machine learning, yet it's one of the most important things to get right. In order to train a predictive model, we need data with a known set of features that needs to be scaled up or down as appropriate. This blog post will explain how feature scaling works and why it's important as well as some tips for getting started with feature scaling. After a scaling operation, the continuous features become similar in terms of range. Although this step isn't required for many algorithms, it's still a good idea to do so. Distance-based algorithms like k-NN and k-Means, on the other hand, require scaled continuous features as model input. There are two common ways for scaling :

Normalization : All values are scaled in a specified range between 0 and 1 via normalisation (or minmax normalisation). This modification has no influence on the feature's distribution, however it does exacerbate the effects of outliers due to

lower standard deviations. As a result, it is advised that outliers be dealt with prior to normalisation. Standardization: Standardization (also known as z-score normalisation) is the process of scaling values while accounting for standard deviation. If the standard deviation of features differs, the range of those features will likewise differ. The effect of outliers in the characteristics is reduced as a result. To arrive at a distribution with a 0 mean and 1 variance, all the data points are subtracted by their mean and the result divided by the distribution's variance.

Python : Python is a widely-used, interpreted, object-oriented, and high-level programming language with dynamic semantics, used for general-purpose programming. It was created by Guido van Rossum, and first released on February 20, 1991. One of the amazing features of Python is the fact that it is actually one person's work. Usually, new programming languages are developed and published by large companies employing lots of professionals, and due to copyright rules, it is very hard to name any of the people involved in the project. Python is an exception. Of course, van Rossum did not develop and evolve all the Python components himself. The speed with which Python has spread around the world is a result of the continuous work of thousands (very often anonymous) programmers, testers, users (many of them aren't IT specialists) and enthusiasts, but it must be said that the very first idea (the seed from which Python sprouted) came to one head – Guido's

Python goals : In 1999, Guido van Rossum defined his goals for Python: an easy and intuitive language just as powerful as those of the major competitors; open source, so anyone can contribute to its development; code that is as understandable as plain English; suitable for everyday tasks, allowing for short development times. About 20 years later, it is clear that all these intentions have been fulfilled. Some sources say that Python is the third-most popular programming language in the world, while others claim it's the fifth. It's used extensively to implement complex Internet services like search engines, cloud storage and tools, social media and so on. Whenever you use any of these services, you are actually very close to Python, although you wouldn't know it. Many developing tools are implemented in Python. More and more everyday use applications are being written in Python. Lots of scientists have abandoned expensive proprietary tools and switched to Python. Lots of IT project testers have started using Python to carry out repeatable test procedures. The list is long. Python is a great choice for: Web and Internet development (e.g., Django and Pyramid frameworks, Flask and Bottle micro-frameworks) Scientific and numeric computing (e.g., SciPy – a collection of packages for the purposes of mathematics, science, and engineering; Ipython – an interactive shell that features editing and recording of work sessions) Education (it's a brilliant language for teaching programming!) Desktop GUIs (e.g., wxWidgets, Kivy, Qt) Software Development (build control, management, and testing – Scons, Buildbot, Apache Gump, Roundup, Trac) Business applications (ERP and e-commerce systems – Odoo, Tryton) Games (e.g., Battlefield series, Sid Meier\'s Civilization IV…), websites and services (e.g., Dropbox, UBER, Pinterest, BuzzFeed…) And that's just the beginning…Python's statements include: The assignment statement, using a single equals sign = The if statement, which conditionally executes a block of code, along with else and elif (a contraction of else-if) The for statement, which iterates over an iterable object, capturing each element to a local variable for use by the attached block The while

statement, which executes a block of code as long as its condition is true The try statement, which allows exceptions raised in its attached code block to be caught and handled by except clauses; it also ensures that clean-up code in a finally block is always run regardless of how the block exits The raise statement, used to raise a specified exception or re-raise a caught exception The class statement, which executes a block of code and attaches its local namespace to a class, for use in object-oriented programming The def statement, which defines a function or method The with statement, which encloses a code block within a context manager (for example, acquiring a lock before it is run, then releasing the lock; or opening and closing a file), allowing resource-acquisition-is-initialization (RAII)-like behavior and replacing a common try/finally idiom[The break statement, which exits a loop The continue statement, which skips the current iteration and continues with the next The del statement, which removes a variable—deleting the reference from the name to the value, and producing an error if the variable is referred to before it is redefined The pass statement, serving as a NOP, syntactically needed to create an empty code block The assert statement, used in debugging to check for conditions that should apply The yield statement, which returns a value from a generator function (and also an operator); used to implement coroutines The return statement, used to return a value from a function The import statement, used to import modules whose functions or variables can be used in the current program The assignment statement (=) binds a name as a reference to a separate, dynamicallyallocated object. Variables may subsequently be rebound at any time to any object. In Python, a variable name is a generic reference holder without a fixed data type; however, it always refers to some object with a type. This is called dynamic typing—in contrast to staticallytyped languages, where each variable may contain only a value of a certain type. Python does not support tail call optimization or first-class continuations, and, according to van Rossum, it never will. However, better support for coroutine-like functionality is provided by extending Python's generators. Before 2.5, generators were lazy iterators; data was passed unidirectionally out of the generator. From Python 2.5 on, it is possible to

pass data back into a generator function; and from version 3.3, it can be passed through multiple stack levels. Python uses duck typing and has typed objects but untyped variable names. Type constraints are not checked at compile time; rather, operations on an object may fail, signifying that it is not of a suitable type. Despite being dynamically-typed, Python is strongly-typed, forbidding operations that are not well-defined (for example, adding a number to a string) rather than silently attempting to make sense of them. Python allows programmers to define their own types using classes, most often used for object-oriented programming. New instances of classes are constructed by calling the class (for example, SpamClass() or EggsClass()), and the classes are instances of the metaclass type (itself an instance of itself), allowing metaprogramming and reflection. Before version 3.0, Python had two kinds of classes: old-style and new-style. The syntax of both is the same, the difference being whether the class object is inherited from, directly or indirectly (all new-style classes inherit from object and are instances of type). In versions of Python 2 from Python 2.2 onwards, both kinds of classes can be used. Old-style classes were eliminated in Python 3.0. The long-term plan is to support gradual typing. From Python 3.5 on, the language's syntax allows specifying static types, but they are not checked in the default implementation, CPython. An experimental optional static type-checker, mypy, supports compile-time type checking

Machine learning algorithms are classified into 4 types:

Supervised Learning

Unsupervised Learning

Semi-supervised Learning

Reinforcement Learning

In a world where nearly all manual tasks are being automated, the definition of manual is changing. Machine Learning algorithms can help computers play chess,

perform surgeries, and get smarter and more personal. We are living in an era of constant technological progress, and looking at how computing has advanced over the years, we can predict what's to come in the days ahead. One of the main features of this revolution that stands out is how computing tools and techniques have been democratized. In the past five years, data scientists have built sophisticated data-crunching machines by seamlessly executing advanced techniques. The results have been astounding. Machine learning algorithms are classified into 4 types: Supervised and Unsupervised Learning Semi-supervised Learning Reinforcement Learning However, these 4 are further classified into more types. What Are The 10 Popular Machine Learning Algorithms? Below is the list of Top 10 commonly used Machine Learning (ML) Algorithms: Linear regression

Logistic regression

Decision tree

SVM algorithm

Naive Bayes algorithm

KNN algorithm

K-means Clustering

Random forest algorithm

Dimensionality reduction algorithms

Gradient boosting algorithm

AdaBoosting algorithm

How Learning These Vital Algorithms Can Enhance Your Skills in Machine Learning :

If you're a data scientist or a machine learning enthusiast, you can use these techniques to create functional Machine Learning projects. There are three types of most popular Machine Learning algorithms, i.e - supervised learning, unsupervised learning, and reinforcement learning. All three techniques are used in this list of 10 common Machine Learning Algorithms.

List of Popular Machine Learning Algorithms

1. Linear Regression : To understand the working functionality of this algorithm, imagine how you would arrange random logs of wood in increasing order of their weight. There is a catch; however – you cannot weigh each log. You have to guess its weight just by looking at the height and girth of the log (visual analysis) and arrange them using a combination of these visible parameters. This is what linear regression in machine learning is like. In this process, a relationship is established between independent and dependent variables by fitting them to a line. This line is known as the regression line and represented by a linear equation Y= a *X + b. In this equation: Y – Dependent Variable a – Slope X – Independent variable b – Intercept The coefficients a & b are derived by minimizing the sum of the squared difference of distance between data points and the regression line.

2. Logistic Regression : Logistic Regression is used to estimate discrete values (usually binary values like 0/1) from a set of independent variables. It helps predict the probability of an event by fitting data to a logit function. It is also called logit regression. These methods listed below are often used to help improve logistic regression models: include interaction terms eliminate features regularize techniques use a non-linear model

3. Decision tree : Decision Tree algorithm in machine learning is one of the most popular algorithm in use today; this is a supervised learning algorithm that is used for classifying problems. It works well classifying for both categorical and continuous dependent variables. In this algorithm, we split the population into two

or more homogeneous sets based on the most significant attributes/ independent variables.

4. SVM (Support Vector Machine) Algorithm : SVM algorithm is a method of classification algorithm in which you plot raw data as points in an n-dimensional space (where n is the number of features you have). The value of each feature is then tied to a particular coordinate, making it easy to classify the data. Lines called classifiers can be used to split the data and plot them on a graph.

5. Naive Bayes Algorithm : A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features are related to each other, a Naive Bayes classifier would consider all of these properties independently when calculating the probability of a particular outcome. A Naive Bayesian model is easy to build and useful for massive datasets. It's simple and is known to outperform even highly sophisticated classification methods.

6. KNN (K- Nearest Neighbors) Algorithm : This algorithm can be applied to both classification and regression problems. Apparently, within the Data Science industry, it's more widely used to solve classification problems. It’s a simple algorithm that stores all available cases and classifies any new cases by taking a majority vote of its k neighbors. The case is then assigned to the class with which it has the most in common. A distance function performs this measurement. KNN can be easily understood by comparing it to real life. For example, if you want information about a person, it makes sense to talk to his or her friends and colleagues! Things to consider before selecting K Nearest Neighbours Algorithm: KNN is computationally expensive Variables should be normalized, or else higher range variables can bias the algorithm Data still needs to be pre-processed.

7. K-Means Clustering : It is an unsupervised learning algorithm that solves clustering problems. Data sets are classified into a particular number of clusters (let's call that number K) in such a way that all the data points within a cluster are

homogenous and heterogeneous from the data in other clusters. How K-means forms clusters: The K-means algorithm picks k number of points, called centroids, for each cluster. Each data point forms a cluster with the closest centroids, i.e., K clusters. It now creates new centroids based on the existing cluster members. With these new centroids, the closest distance for each data point is determined. This process is repeated until the centroids do not change.

8. Random Forest Algorithm A collective of decision trees is called a Random Forest. To classify a new object based on its attributes, each tree is classified, and the tree “votes” for that class. The forest chooses the classification having the most votes (over all the trees in the forest). Each tree is planted & grown .If the number of cases in the training set is N, then a sample of N cases is taken at random. This sample will be the training set for growing the tree.

9. Dimensionality Reduction Algorithms : In today's world, vast amounts of data are being stored and analyzed by corporates, government agencies, and research organizations. As a data scientist, you know that this raw data contains a lot of information - the challenge is in identifying significant patterns and variables. Dimensionality reduction algorithms like Decision Tree, Factor Analysis, Missing Value Ratio, and Random Forest can help you find relevant details.

10. Gradient Boosting Algorithm and AdaBoosting Algorithm These are boosting algorithms used when massive loads of data have to be handled to make predictions with high accuracy. Boosting is an ensemble learning algorithm that combines the predictive power of several base estimators to improve robustness. In short, it combines multiple weak or average predictors to build a strong predictor. These boosting algorithms always work well in data science competitions like Kaggle, AV Hackathon, CrowdAnalytix. These are the most preferred machine learning algorithms today. Use them, along with Python and R Codes, to achieve accurate outcomes.

KEY LEARNING FROM INTERNSHIP :

An internship is an opportunity to test drive a career without making any serious commitments. It provides you with experiences, lessons, and the tools you'll need to get a full-time gig in the future.

It is often a great choice because it gives you a feel for work without being thrown into the deep end straight away. This provides you with the opportunity to grow and learn before fully entering the working world.

Here are all the things that I gain from an internship :

- **New and improved skills and how to apply them :**

One of the most important things you can gain from an internship is newfound knowledge. This can include knowing how to fulfil tasks that are relevant to your desired career path and sharpening the skills that you already possess.

Many students think an internship mainly consists of making coffee and running errands for superiors all day, but that is not true. And really, it shouldn't be like that. An internship is an opportunity to test out all the skills that you developed in varsity/college and see how they work in the real world.

With internship opportunities, such as the Doctors Without Borders organisation, you will get an idea of what your biggest strengths are, as well as areas of improvement you should work on.

- **Professional communications :**

Working in a professional setting for the first time can be difficult to get used to. But it is the best way to learn how to navigate the working world through real-life, hands-on experience.

One of the most valuable skills you will gain from an internship is the ability to speak with people in a professional setting. Discussions with bosses or coworkers are different from discussions with lecturers or fellow students,

After your internship, you should have a better idea of the appropriate way to behave as a professional. This will help you a lot when you start interviewing for jobs because you will be more confident and will sound more mature and experienced in a business setting.

- **Networking is important :**

Don't underestimate yourself; make sure you make the most of your internship and take advantage of all the opportunities that come with it.

Also, unglue yourself from your desk every once in awhile and get to know other interns. Not only will you end up creating great memories and making friends, but you will widen your professional network.

- **Taking constructive criticism well :**

Naturally, no one likes to be criticised and performance evaluations can be quite scary. You will probably make a few mistakes and receive constructive criticism about your work from both your colleagues and your boss.

Always remind yourself that it's not personal. It is for your own good and growth and it will improve the quality of your work.

- **Work hard no matter what you're doing :**

Always work hard even if your task is small and seems unimportant. It will help you build a good work ethic, and people will notice the effort you put in.

It's not nice being told what to do all the time, but your superiors (mostly) know better. Following the rules and instructions they give you makes it easier for everyone.

- **Independence :**

Often, we think being spoon-fed is the way to learn, but working independently has proved to be very important. Your internship will teach you to make my own decisions and do things on your own.

Being able to work independently with little guidance is very important in the working world.

- **Making connections :**

In addition to the people who will be your references in the future, try to leave your internship with new connections: senior employees, clients, fellow interns, etc.

These people can provide guidance, advice and help you in future job searches. Keep them in the loop on where you are in your career, and offer to help them whenever you can.

- **You're more important than you think :**

Even though you are at the bottom of the career ladder, you're also needed and your basic work is appreciated.

Intern life can be tough, and you only get a short time to make a lasting impression. However, it is a great opportunity to gain experience, make friendships and learn. So work hard and enjoy your internship!

INTERNSHIP / PROJECT DISCUSSION :

- **Brief Objectives of Project –**

In this Internship we have to build an AI Enabled Fintech B2B Invoice Management Application.

Introduction to B2B Operations :

The B2B world operates differently from the B2C or C2C world. Businesses work with other businesses on credit. When a buyer business orders goods from the seller business, the seller business issues an invoice for the same. This invoice for the goods contains various information like the details of the goods purchased and when it should be paid. This is known in accounting terminology as “Accounts Receivable”. “Accounts Receivable represents money owed by entities to the firm on the sale of products or services on credit. In most business entities, accounts receivable is typically executed by generating an invoice and either mailing or electronically delivering it to the customer, who, in turn, must pay it within an established timeframe, called credit terms or payment terms.”Seller business interacts with various businesses and sells goods to all of them at various times. Hence, the seller business needs to keep track of the total amount it owes from all the buyers. This involves keeping track of all invoices from all the buyers. Each invoice will have various important fields like a payment due date, invoice date, invoice amount, baseline date etc.

The buyer business needs to clear its amount due before the due date. However, in real-world scenarios, the invoices are not always cleared i.e. paid in full amount by the due date. The date on which a customer clears the payment for an invoice is called the payment date.

Account receivables Department :

1. In the ideal world, the buyer business should pay back within the stipulated time (i.e. the Payment Term). However, in the real world, the buyer business seldom pays within their established time frame, and this is where the Account Receivables Department comes into the picture.

2. Every business consists of a dedicated Account receivables Department to collect and track payment of invoices.

3. It consists of an Account receivables team that is responsible for :

- Collecting payments from customers for their past due to invoices.
- Sending reminders and follow-ups to the customers for payments to be made.
- Looking after the entire process of getting the cash inflow.

Help the company get paid for the services and products supplied.

About Machine Learning Part :

As a winter internship project, you will be building a web application to help the people working in the Accounts Receivable departments in their day-to-day activities. You need to build a web application where the users in the Account Receivable department can :

- View the invoice data from various buyers.
- See various fields/attributes of the invoice(s) from a particular buyer.
- Perform Data Pre-processing on the invoice data.
- Get account-level analytics to easily visualize and interpret data- EDA and Feature Engineering.
- Get a prediction of when the invoice is going to get paid.

About Web Application Development Part :

The objective of the Web Application Development internship project is :

- To build a Full-stack Invoice Management Application using ReactJs, JDBC, Java, Servlets.
- Build a responsive Receivables Dashboard.
- Visualize Data in the form of grids.
- Visualize Data in the form of graphs.
- Perform Searching operations on the invoices.
- Add & Edit data in the editable fields of the grid.
- Delete data of selected rows in predefined templates.

About React Web App Part :

The mandatory features are compulsory tasks and the optional features are for extra credit points, which will give you an added advantage.

Mandatory Features	Optional Features
1. UI Creation 2. Grid Creation 3. Grid Data Loading 4. Crud Operation a) Add b) Edit c) Delete 5. Pagination 6. Advanced Search	1. Predict Button activation with Grid Data 2. Shortcut search button on Grid for Customer Id 3. Sorting columns 4. View - Analytics

- **How the Objectives were achieved -**

This Internship is of 11 weeks in which we have to build an AI Enabled Fintech B2B Invoice Management Application. Basically the whole task is divided into 5 parts :

- Machine Learning
- Java
- HTML & CSS
- JavaScript(React JS)
- Web App Project

Firstly I have completed the Machine Learning and Java part. As here i will be building a web application to help the people working in the Accounts Receivable departments in their day-to-day activities. In this web application users in the Account Receivable department can View the invoice data from various buyers. He can see various fields/attributes of the invoice(s) from a particular buyer. Also he can perform Data Pre-processing on the invoice data.

He will get account-level analytics to easily visualize and interpret data- EDA and Feature Engineering. and he get a prediction of when the invoice is going to get paid. So for these above tasks I have created a ML model using a dataset (which is provided by HighRadius Corporation) and other ML algorithms. Following are the steps for model preparation :

Data-Cleaning and Preprocessing : Following are the steps that we will follow to prepare the dataset the training of ML model -

- As we can see in the dataset Image, the format the dates are of different timestamps.
- The data also includes the NULL values we will replace the mean values

with median.

- We will convert the categorical data like (USD, Rupee) to the numerical format like 0,1 etc
- As the date includes the information like month, year and day so we will separate the values accordingly.
- We will calculate the difference between the columns Clear_date, due_term values that will be output of our Model.

Exploratory Data Analysis : Following are the steps that we will follow -

- The date includes some unnecessary information like S. No. which does not give any information about data.
- We will remove those columns which do not plays that much role for the prediction using the correlation Matrix.
- We will remove the outliers from the data and normalize it so that the data will be balanced.

Model Preparation : After cleaning and balancing the data we will split that data in ration 8 : 2 for the training and test. For the prediction we will use Machine learning algorithms like SVM, Random Forest etc. As the Model will be trained, we'll save the model file to connect it with the backend of the system.

For Backend in my web application I have used Java. Following are the functionalities that I'll be providing in the Backend of the system. For the prediction of Model, we need the data so we will be creating the database whose architecture is as following-

- **All the Columns of the CSV file need to be loaded into the DB.**

List of all the fields part of dataset are as follows:

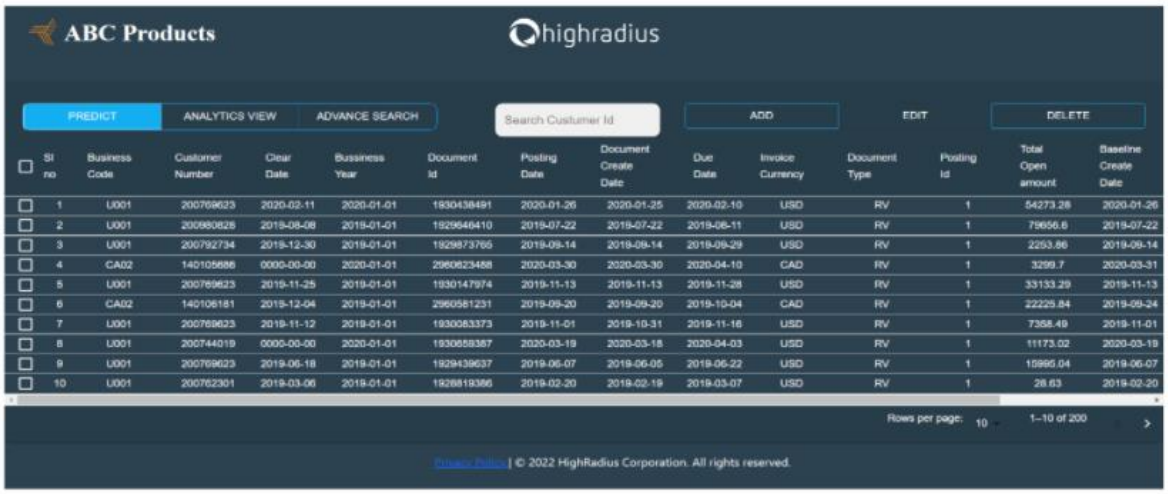
- sl_no
- business_code
- business_name
- cust_number
- name_customer
- clear_date
- buisness_year
- doc_id
- posting_date
- document_create_date
- document_create_date.1
- due_in_date
- invoice_currency
- document type
- posting_id
- area_business
- total_open_amount
- baseline_create_date
- cust_payment_terms
- invoice_id
- isOpen
- predicted

These column values will be added in the database column so that we can implement ML Prediction on Real life data. Using the Servlet and JSP we will connect the Database and Frontend elements. We will add the CRUD functionalities so that the user can manage the data accordingly. We will also add the search functionality as we will be connecting the database .The dataset that I am using for this project is as following -

- business_code: company code of the account
- cust_number: customer number given to all the customers of the account
- name_customer: name of the customer
- clear_date: date on which the customer clears an invoice, or in simple terms, makes the full payment
- business_year: the year in which the invoice was created
- posting_date: the date on which the particular invoice was entered in the ERP database
- document_create_date: the date on which the invoice document was created
- document_create_date.1: normalized version of document_create_date (we'll use this to split the data)
- due_in_date: the date on which the customer is expected to clear an invoice

- invoice_currency: the currency of the invoice amount in the document of the invoice
- document type: represents the type of document, eg, D1 represents invoice
- posting_id: key indicator to identify whether on AR terms is invoice, deduction, credit memo basedon its value. Applicable for SAP ERP
- area_business: business area in sap is defined as an organisational area within the financialaccounting module
- baseline_create_date: the date on which the invoice was created and accounts ondiscounts and days of payments
- invoice_id: unique number assigned when a seller creates an invoice

After that I design the UI for the project. The payment date needs to be persisted across sessions in the UI. The UI for the project consists of a single screen which is as follow –

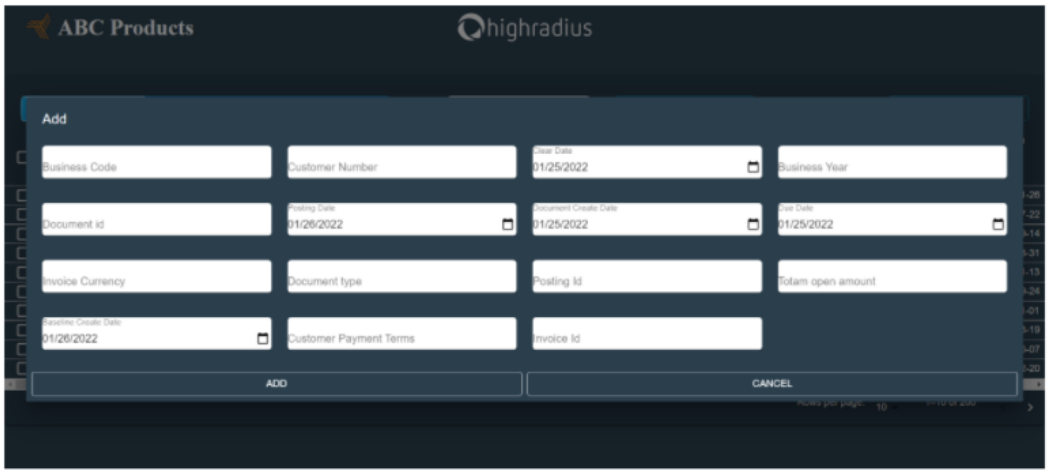


a.) First Section is the header which comprises the ABC Product logo on the left, the HighRadius Logo in the middle.

b.) The second section consists of Predict, Advance Search, Analytics View Add, Delete & Edit, and Search bar.

Add button :

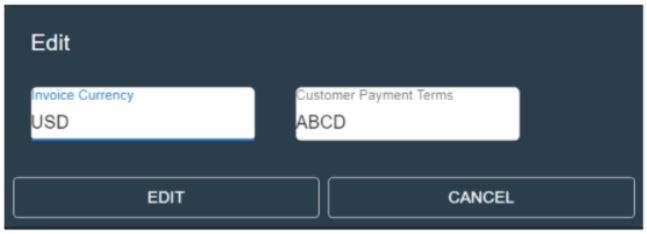
It is used for adding new field values to the grid. The Add button will be in the enabled state if no row is selected. Whenever one or more rows are selected, the Add button will still remain activated. After clicking on the Add button, a pop-up window will appear with all the fields for which values need to be added along with a Cancel and an Add button. The user should be able to type in the values, except for the date of the invoice for which there should be a calendar view from where the user is able to select the required date, month, and year. The user should fill in all the required fields before adding. If the user tries to click on add before all mandatory fields are filled, the user will not be able to add.

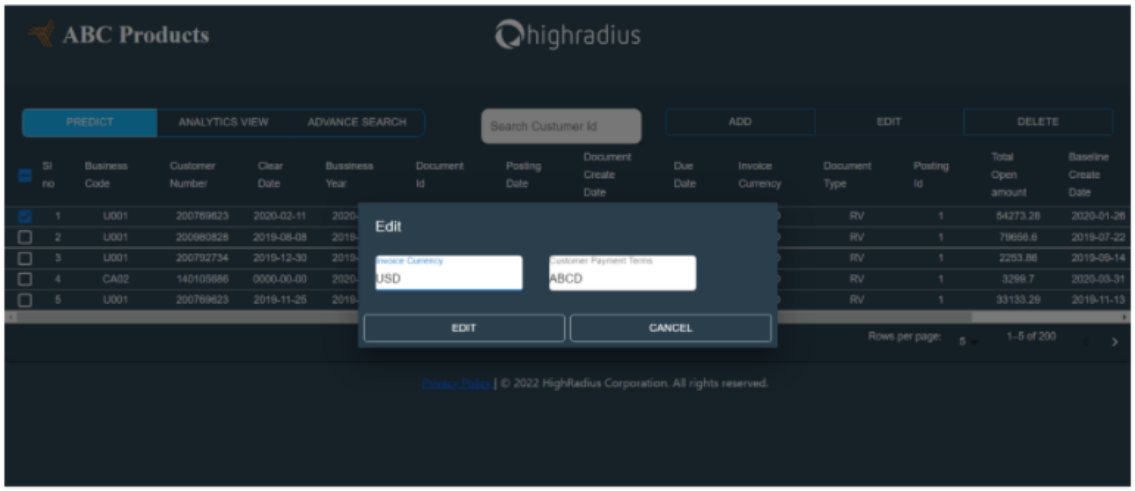


FULL SCREEN VIEW

Edit button :

It is used for editing the editable field values in the grid. Edit button should be disabled at first and should enable only one checkbox is Selected . A user should be able to select a row and then click on the Edit button. The fields which can be edited are the Invoice Currency and Customer Payment Terms fields. Without selecting any row, the Edit button should remain disabled. On clicking the Edit button, a popup should open up with the column header name and existing value. The user should be able to edit the existing value. The popup should have a Edit, Cancel as shown in the UI below :

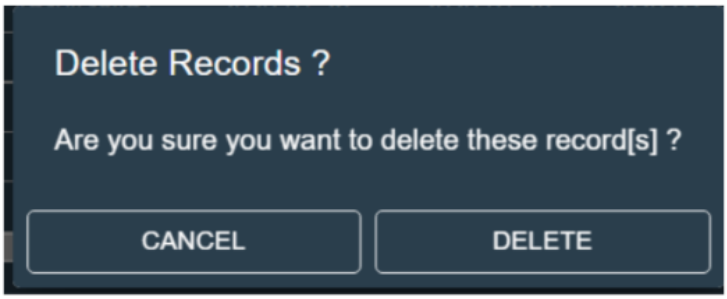


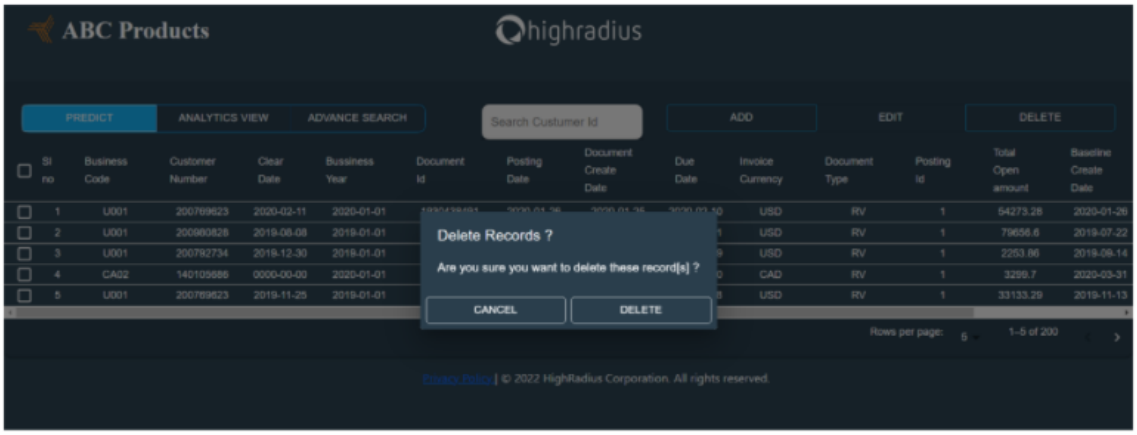


FULL SCREEN VIEW

Delete Button :

Clicking on the delete button will allow the user to delete records from the grid. When the user selects one or more rows, the delete button gets enabled. A pop-up should be displayed on clicking delete to confirm that the user wants to delete the selected records permanently. Once the user clicks on the delete button, the row(s) should be removed from the grid in the UI and should remain persistent.



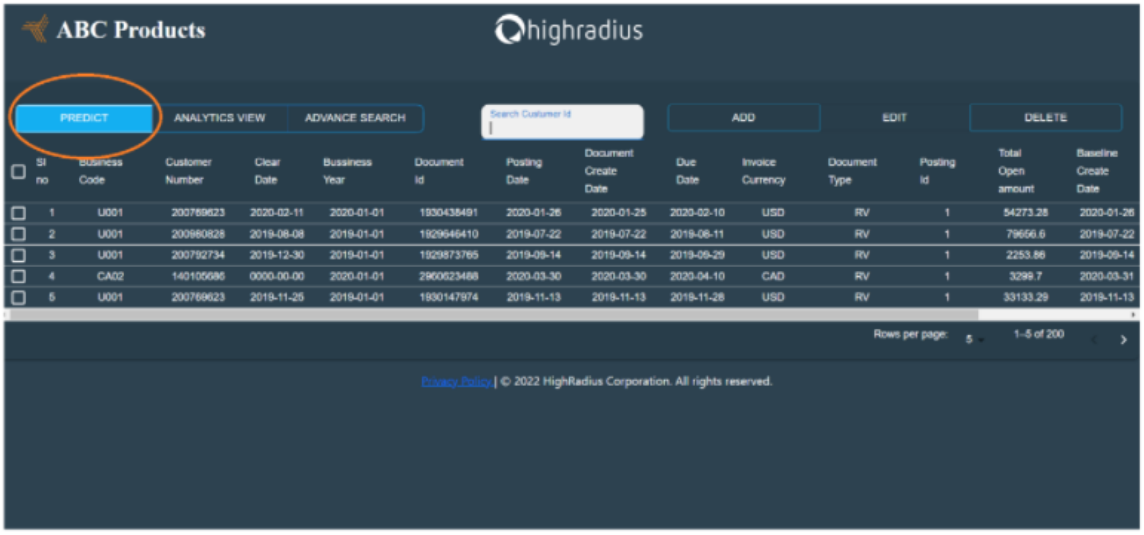


FULL SCREEN VIEW

Predict button :

Users should be able to predict the payment date of selected Invoices with the help of the Predict button. Clicking on this button will populate the Predicted Payment Date column on the UI with the predicted dates. When the user selects one or more

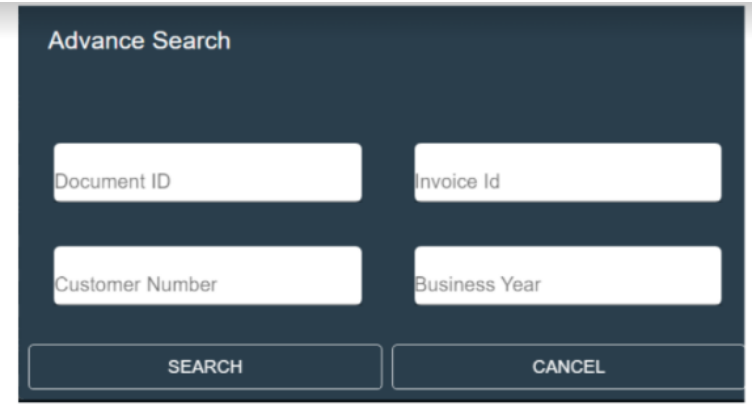
Invoices and clicks on the Predict button, the Predicted Payment Date column should get populated only for those invoices. The button should get activated only upon selecting any of the Invoice(s). If no Invoice is selected, the button should be in a disabled state.



Advanced Search button :

The UI consists of the Advanced Search button. Clicking on this button will help the user to perform an advanced search on the data based on the following four fields -

- Document Id-(doc_id)
- Customer No-(cust_number)
- Invoice No-(invoice_id)
- Business Year- (buisness_year)



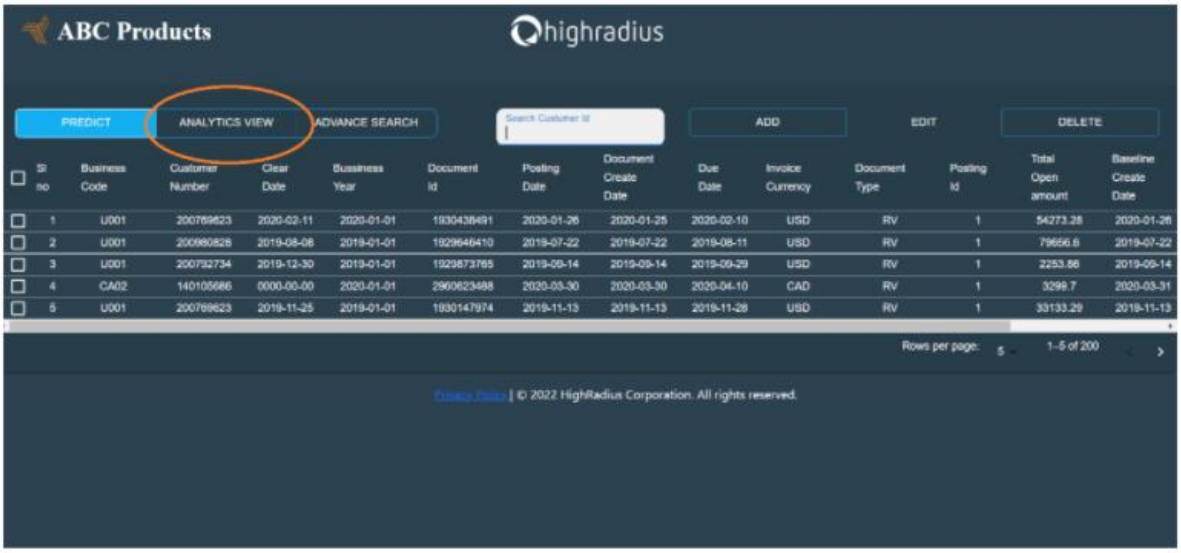
FULL SCREEN VIEW

Analytics View :

To get insights from the existing data based on users inputs. The existing parameters would act as key points or outliers for the synthesis of data. So the analytics view will be a button in UI which responds to a new window on click event. The new window contains of parameters:

- Currency (Multiselect)
- Due Date
- Baseline Create Date
- Clear Date

The user will have a privilege to go for single parameter or multi parameter based on their choices and preferences. On submitting the parameters the web application will open the dialog window which will provide the user with an illustration of a bar graph and pie chart which will be formed based on the parameterized data that the user had selected. The bar graph will be showing data for the total open amount and number of customers for all Business. There will be a close button to close the window and redirect the user to the main screen(UI). The analytics view button would be a simple tool that the user can use to view data based on their preferences and could facilitate decision making.



Footer :

The Grid consists of the Footer which will have the following three parts :

● 'Viewing <starting count> - < end count> of <total count>' text on left. It shows invoices currently active.

Example1: Viewing 1-10 of 500 means that the user is seeing 1-10 Invoices present on the page out of the total number of invoices which is 500.

● Pagination arrows with text ' <present page number> of <total page number> ' in center. Clicking on the back arrow takes the user to the previous page. Clicking on

the next arrow takes the user to the next page. Back arrow should be disabled when the user is on the first page and the next arrow should remain disabled if the user is on the last page.

Example2: 2 of 50 means that the user is currently on 2nd page and seeing invoices 11-20 out of the total 500 invoices.

● 'Copyright 2022 Highradius.All Rights Reserved.' in the middle.

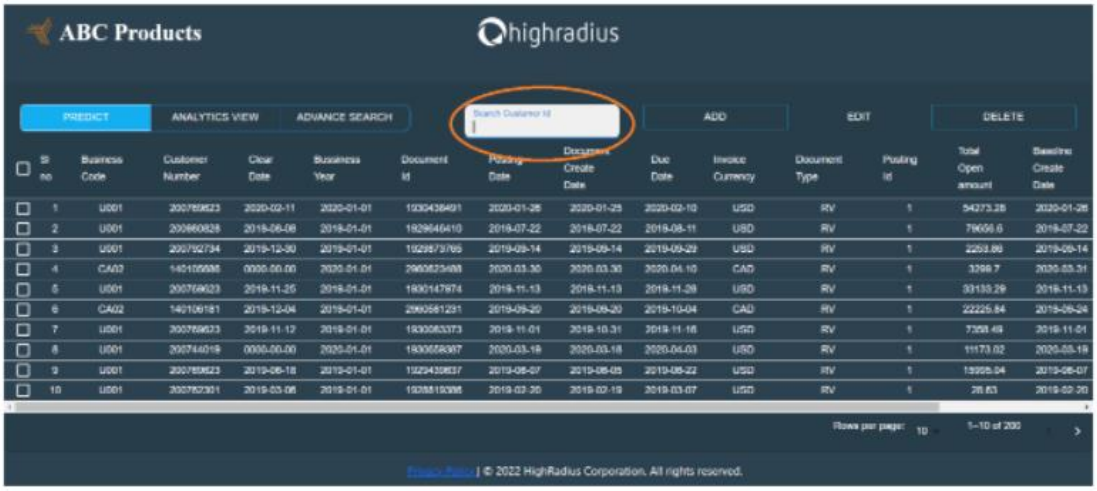
After the creation of Add, Edit, Delete, Predict , Advance Search ,Analytic View and Footer I have created the next remaining UI and also give the functionalities to the created buttons. After this I have implemented search functionality. After that I have implemented Grid Data Analytics , Data Storing and Pagination.

Searchable fields –

We can perform search functionality on the basis of following data :

- Business Year- Text Field
- Customer Id-Text Field
- Invoice No - Text Field
- Document Id - Text Field
- Customer Id - Equal Search

Users should be able to search for a customer by typing text in the Customer id integer field. Search is not case-sensitive.



Grid Panel Section:

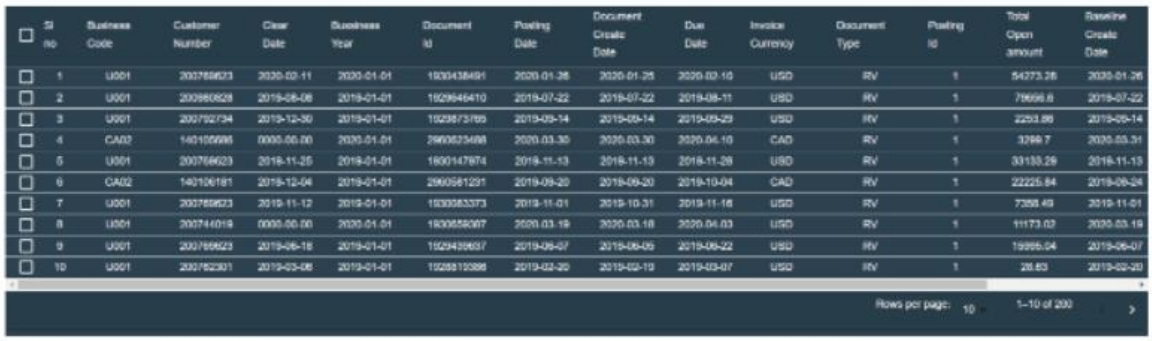
The Seventh Section is the Grid Header section, consisting of all the different column name headers and a Select All and Deselect All functionality. Following are the columns to be displayed in the UI:

- sl_no
- business_code
- cust_number
- clear_date
- buisness_year
- doc_id
- posting_date
- document_create_date
- due_in_date
- invoice_currency
- document type
- posting_id
- total_open_amount
- baseline_create_date
- cust_payment_terms

- invoice_id

The Grid panel section will be divided into 3 portions:

- The header of the grid will have a Predict button on the top left corner followed by an Advance Search Button,an Analytics view, an Add Button, an Edit Button, a Delete Button, and a Search Bar.
- The name of the grid that is Invoice List will be mentioned in the top left corner of the grid.
- The second portion is the table with customer invoice data as rows and the colums.



The Grid consists of the Grid Rows that contains the required data that is loaded from the CSV File. On a single page, only 10 invoices' data is displayed. Users can select single or multiple rows.

Sorting & Searching :

1. Sorting: Sorting should be performed on all the columns:

2. Clicking on the column headers should sort the values of the whole grid

- First click – Ascending
- Second click - Descending

Column headers should have a double arrow symbol near the column name to

indicate they can be sorted.

Horizontal Scroll Bar : The Horizontal Scroll Bar which can be used to scroll across the screen to access the data in various columns.

AI support for the prediction of payment date :

- As part of this project, you need to predict the Payment Date of each invoice.
- In order to achieve this, there should be a button named “Predict” present on the UI besides the “Advance Search” button.
- Users can select one or more invoices and click on the “Predict” button to predict the payment dates of those selected invoices.
- Once the button is clicked, the Predicted Payment Date column should get populated with the predicted dates derived from the ML model.

The “Predict” button should remain disabled if no invoices are selected.

- **What skills (scientific and professional) were learned during the internship?**

There are many skills that you learn with the help of internship or training. Some of the skills are as following -

- Communication Skills -

There are various ways to communicate these days, but potential employers tend to look towards your ability to speak and write professionally. You must make the best of every opportunity to demonstrate your verbal skills & writing prowess. Your resume & cover letter must be engaging in well-written format, and you should also be able to supply thoughtful answers to likely interview questions. Your ability to communicate effectively by conveying information and translating ideas among your coworkers, supervisors, or clients, becomes essential in any field.

- **Experience of Work -**

Yes, we all need work experience before we can be considered as employable but don't be deceived, it isn't quite the same as this next skill. It's something we tend to overlook when taking on an internship role but, getting an experience of your career of choice and the hands-on knowledge that can only be acquired by actually working at a business are two very different things. Experience of work will involve you learning more than you have ever known about workplace politics and office relationships; which brings us to our next skill.

- **Teamwork Skills -**

Internships are professional roles where every team member must work together to accomplish any task. During internships, you'll have to remain focused on the

team's goals. You should also develop a patient, positive, and considerate attitude whenever disagreement arises among members of your team.

- **Software Education -**

As our world becomes increasingly reliant on technology, we must look at every opportunity to improve our software skills to become more capable future employees. Internships are an excellent opportunity for all-round improvement. Apart from making your resume look good, they have so much more to give. By learning to use software like Adobe Photoshop, Acrobat, or In Design, during an advertising internship, you can deliver more value at a future job. Moreover, because many software skills are transferable across different industries, you can switch professions and remain relevant.

- **Critical Thinking & Problem Solving Skills -**

Here, we are referring to your overall problem-solving skills; your ability to evaluate and analyze an issue or a situation and then respond effectively. This can be demonstrated by your enthusiasm to ask questions and examine issues from various angles. Internship roles will help you think critically to provide creative solutions to work challenges. This is a skill that is highly valued and emphasized by many by employers.

- **Project Management Skills -**

You must learn how to cope with workload early in your career. Internships offer excellent learning experiences in this regard. You'll learn to take responsibility and prioritize your career goals. And if you haven't developed adequate project management skills, your internship supervisor will likely give you a few pointers on the best ways to resolve work problems. These solutions may involve you getting a work planner to map out your daily goals and targets so you and your team members can be aware of them. On the other hand, at non-internship, full-time job roles, your manager probably won't have the time to mentor you hence

learning how to properly manage work projects during internships becomes crucial for your future performance.

- **Networking -**

Proper networking is another valuable skill you can learn during your internship. While you may most likely be working as part of a departmental team, exploring outside the box options is also a great experience that will certainly improve your overall communication with others.

- **Results/observations/work experiences get in the internship company –**

Internships offer a wealth of opportunities and experiences early in your career. Being an intern provides you with a chance to practice what you've learned in a physical or virtual work environment.

Virtual internships are not a new concept. But technology accessibility, coupled with the COVID-19 pandemic, make them a suitable option nowadays. Effective programs combine learning, training, mentoring, and networking.

Here are a few benefits you'll likely encounter during your virtual internship experience.

- **Take Advantage of Flexibility**

Virtual internships solve scheduling issues facing students. Calendars quickly fill up with classes, extra-curricular activities, family obligations, and jobs. Many college students juggle employment and studies at the same time.

The National Center for Education Statistics notes that in 2018, 43 per cent of full-time undergraduate students worked while enrolled.

When you intern remotely, you can often work at your convenience. You just need to log the requisite number of hours per day or for the project.

Instead of a strict Monday-to-Friday schedule, you can choose to utilise your free time and weekends for the internship.

- **Practice Your Communication Skills**

Scheduling conflicts and lack of in-person interaction emphasise the need for regular and effective communication.

The virtual environment is very different from the face-to-face environment. So, you will have to master the art of conversation and negotiation digitally.

Emails: How do you put your intentions into words? How do you follow up on requests firmly and politely? How do you write a professional response when you are angry?

Video conferencing: How do you remain focused while others are speaking? How do you voice out your concerns and suggestions? How do you appear professional on video?

- **Widen Your World and Connections**

If you want to explore international companies or organisations, a virtual internship is a great option. It eliminates the usual concerns, such as commuting and limiting your choices because of location.

Aside from work itself, you can build relationships with people from diverse backgrounds and cultures.

Establishing networks is not just for business-minded individuals but for anyone keen on building a successful career.

People you meet at work and school can help you out in the future, and vice versa.

- **Prepare for the Future of Work**

It appears the option to work remotely is here to stay. Due to the necessity amid the pandemic, businesses have recognised that the arrangement can work.

Twitter, for one, has expressed openness to employees wanting to work from home permanently.

Your virtual internship prepares for telecommuting and other flexible working arrangements in the future.

Being able to work with technology, as well as tools for project management and productivity, is a must these days.

Plus, you'll already know what to do if technical issues threaten to disrupt your workflow.

- **Develop How You Work**

The flexibility afforded by a virtual internship is not without responsibility. You still need to report to your supervisor, meet project deadlines, collaborate with others, and remain productive while working independently.

Accountability, self-discipline, time management, patience, resourcefulness, and proactiveness are among the values needed for this setup to work. These will also become the foundation of your work ethics moving forward.

- **Discover Where You Want to Go**

Interning helps you figure out a lot of things. Do you want to pursue a career in your field of study, develop a specialisation, or explore other paths? Are you productive working from home, or do you like the energy of working in an office?

From business to fashion, virtual internships give you a taste of a practitioner's work. As an intern, you will experience first-hand the exciting and mundane aspects of the profession.

And work with other teams or departments may spark newfound interests.

- **Enhance Your Employability**

As designed, a virtual internship lets you apply what you learned in school to a workplace setting. It constitutes experience, a premium when you look for employment after graduation.

For some colleges, internships are integrated into the curriculum and are required.

Besides experiences, you gain skills and develop qualities from being an intern. Transferable skills, or soft skills, are advantageous because they are relevant regardless of the job, employer, or industry.

You'll grow in teamwork, leadership, communication, listening, decision making, and problem-solving. And that's just to name a few.

Work anytime, anywhere. Discover your passion and connect with people from all over the world.
At the same time, increase your employability and future success. For your resume and personal goals, a virtual internship is worth pursuing.

- **What challenges did you experience during the internship?**

In Machine Learning , there occurs a process of analyzing data for building or training models. It is just everywhere; from Amazon product recommendations to self driven cars, it beholds great value throughout. As per the latest research, the global machine learning market is expected to grow by 43% by 2024. This revolution has enhanced the demand for machine learning professionals to a great extent. AI and machine learning jobs have observed a significant growth rate of 75% in the past four years, and the industry is growing continuously. A career in the Machine learning domain offers job satisfaction, excellent growth, insanely high salary, but it is a complex and challenging process. There are a lot of challenges that machine learning professionals face to inculcate ML skills and create an application from scratch. The challenges I faced during the creation of Machine Learning Model for this project is as following –

- **Poor Quality of Data :**

Data plays a significant role in the machine learning process. One of the significant issues that machine learning professionals face is the absence of good quality data. Unclean and noisy data can make the whole process extremely exhausting. We don't want our algorithm to make inaccurate or faulty predictions. Hence the

quality of data is essential to enhance the output. Therefore, we need to ensure that the process of data preprocessing which includes removing outliers, filtering missing values, and removing unwanted features, is done with the utmost level of perfection.

- **Under-fitting of Training Data :**

This process occurs when data is unable to establish an accurate relationship between input and output variables. It simply means trying to fit in undersized jeans. It signifies the data is too simple to establish a precise relationship. To overcome this issue we Maximize the training time, Enhance the complexity of the model , Add more features to the data, Reduce regular parameters and Increase the training time of model.

- **Over-fitting of Training Data :**

Over-fitting refers to a machine learning model trained with a massive amount of data that negatively affect its performance. It is like trying to fit in Oversized jeans. Unfortunately, this is one of the significant issues faced by machine learning professionals. This means that the algorithm is trained with noisy and biased data, which will affect its overall performance. Let's understand this with the help of an example. Let's consider a model trained to differentiate between a cat, a rabbit, a dog, and a tiger. The training data contains 1000 cats, 1000 dogs, 1000 tigers, and 4000 Rabbits. Then there is a considerable probability that it will identify the cat as a rabbit. In this example, we had a vast amount of data, but it was biased; hence the prediction was negatively affected.

- **Machine Learning is a Complex Process :**

The machine learning industry is young and is continuously changing. Rapid hit and trial experiments are being carried on. The process is transforming, and hence there are high chances of error which makes the learning complex. It includes

analyzing the data, removing data bias, training data, applying complex mathematical calculations, and a lot more. Hence it is a really complicated process which is another big challenge for Machine learning professionals.

- **Lack of Training Data :**

The most important task you need to do in the machine learning process is to train the data to achieve an accurate output. Less amount training data will produce inaccurate or too biased predictions. Let us understand this with the help of an example.

Consider a machine learning algorithm similar to training a child. One day you decided to explain to a child how to distinguish between an apple and a watermelon. You will take an apple and a watermelon and show him the difference between both based on their color, shape, and taste. In this way, soon, he will attain perfection in differentiating between the two. But on the other hand, a machine-learning algorithm needs a lot of data to distinguish. For complex problems, it may even require millions of data to be trained. Therefore we need to ensure that Machine learning algorithms are trained with sufficient amounts of data.

- **Slow Implementation :**

This is one of the common issues faced by machine learning professionals. The machine learning models are highly efficient in providing accurate results, but it takes a tremendous amount of time.

Slow programs, data overload, and excessive requirements usually take a lot of time to provide accurate results. Further, it requires constant monitoring and maintenance to deliver the best output.

- **Imperfections in the Algorithm When Data Grows :**

So you have found quality data, trained it amazingly, and the predictions are really concise and accurate. Yay, you have learned how to create a machine learning algorithm!! But wait, there is a twist; the model may become useless in the future as data grows.

The best model of the present may become inaccurate in the coming Future and require further rearrangement. So you need regular monitoring and maintenance to keep the algorithm working. This is one of the most exhausting issues faced by machine learning professionals.

For Backend Development of my project I have faced many challenges like Selecting an IDE for Java App Development , Memory Leaks , Caching, Database maintenance , Synchronization, Testing.

The major challenges faced during frontend designing part of this project is as follow –

- **Choosing the right tech stack :**

In essence, they're the set of technologies a web development and design agency uses to build a web or mobile app development. You should always choose a tech stack that aligns with the problems you are trying to solve. For example, you probably won't need a complex tech stack for a simple web application, or a tech stack that helps you optimize for scalability when your user base will be a consistent size. You should consider whether your tech stack is widely used in the industry. With industry-standard tech stacks, you will have a large pool of skilled developers to draw from for your initial build and future requirements. Make sure

to choose a tech stack that has thorough documentation.. Great documentation and support can save you time, money and hassle.

- **UX :**

User experience (UX) encapsulates the reactions, perceptions, and feelings your users experience while engaged in your application. It's the feeling of ease and simplicity that you get from great design. It's also the frustration that you feel when interacting with poor design. That's why it's important to think about the overall impression you want to leave on your users before you start making detailed decisions about how to build it. If you deliver effectively to the deeper needs of all your users – providing pleasure, engagement, and an overall emotional appeal – your application can become a central part of your users' day-to-day lives and deliver to your business goals.

- **UI & Simplistic Design :**

User Interface (UI) design includes all the visual elements your users interact with on your web application. It is everything that your users see on their screens and everything they click on to guide them through the experience.Great UI design certainly makes your application visually appealing, but it goes beyond just simple aesthetics. The goal is to make the actual user experience simple and accessible – and usable. This means using only a targeted, purposeful selection of copy and content, making clear the options your users have throughout the experience and ensuring information is readily available at each step.

Intuitive UI typically involves :

- Clear navigation
- Engaging visuals
- Easy-to-read typography

- **Performance and Speed :**

No user likes slow load times. And they can have real consequences for your business. If your application is slow, users won't wait. They'll leave. This is the reality of web application development issues today. You may only get a single chance to hook a user on your product.

So, if you know that you're building an application with a lot of content (e.g. videos), you should outline that upfront so your developers can build a more robust application to ensure performance. You don't want your application to slow down if it makes a splash in the market or you see periodic surges in traffic.

This kind of planning for the future will ensure your initial app launch can deliver the kind of speed and performance you want for your early, often-critical, users. It will also mean that your initial build will provide the foundation you need for future business growth.

- **Scalability :**

The challenge of scalability relates to how you want your application to develop over time. If you want your application built right today, you'll need to know as much as possible about what you need it to do in the future.

Your application might include fairly lean content at launch but, a year or two in, you could be planning a detailed, expansive content-rich experience. This is where extensibility comes in. Extensibility is when an application is initially designed to incorporate new capabilities and functionality in the future. This can be integrated right from the start of the development process. If you can articulate a long-term vision for your app, it can be built to grow and evolve over time.

Planning for scalability helps you manage different user types, handle increased traffic, and an expansion of, for example, an ecommerce shop. Overall, it's

valuable to prioritize scalability because it can improve your user experience, fulfill your business goals and extend the lifespan of your application.

- **Web Security Threats :**

There are a number of things to prioritize so your application and your users are secure. Choosing the right development infrastructure is one. Make sure that the infrastructure you build on has enough security services and options so your developers can implement the proper security measures for your application. SSL certificates are a global standard security technology that enables encrypted communication between your web browser and server. When integrated in your app, they enhance its security and eliminate the chance it is flagged as unsecure by web browsers. SSL certificates also help protect credit-card numbers in ecommerce transactions and other sensitive user information like usernames, passwords and email addresses. In essence, SSL allows for a private “conversation” just between the two intended parties and hides sensitive information from hackers and identity thieves. Creating robust password requirements and multi-factor authentication for your users are also effective security measures. More complex passwords are less likely to be hacked. Multi-factor authentication, where your users take multiple steps to confirm their identities, also gives your application an added layer of protection.

CONCLUSION :

In a nutshell, this internship has been an excellent and rewarding experience. I can conclude that there have been a lot I've learnt from my work at Highradius Corporation . Needless to say, the technical aspects of the work I've done are not flawless and could be improved provided enough time. I believe my time spent in research and discovering it was well worth it and contributed to finding an acceptable solution to build a fully functional web service. The key contribution of this internship is that it not only provides an updated view of research methods in the field of Machine Learning but also discusses emerging issues that are worth the attention of researchers in the field of machine learning. The findings of the study should help Machine Learning researchers to make informed decisions while pursing research in machine learning. This internship has focused more on general research methods in machine learning. Future researchers can perform a content analysis of specific areas of machine learning such as supervised learning, unsupervised learning, video analytics, text analytics , classification, and prediction. Two main things that I've learned the importance of are time-management skills and self-motivation. This internship enhances my machine learning, java and web-app development skills. The experience you gain during an internship increases your market value. You gain skills and credibility on an internship, as well as essentially moving from someone with no professional experience to a candidate with some experience. The distinction might seem small, but it's powerful, and a collection of various internships can do wonders for your personal brand. Internships can help you make up your mind about what you want to do. Although this might include convincing you to adopt a complete change of career, internships can often help you consider which part of an industry to specialize in. Specializations are often lucrative and rewarding career avenues where you will be immeasurably valued for a deep understanding of a certain skill or sector, rather than just having generalist knowledge across lots of different responsibilities.

THANK YOU