

Employee Performance and CRM Analysis of UCI Dataset using Machine Learning Approach

Aishwarya T¹, Shalini S², Julia Faith S³

^{1,2}UG Student, Department of IT, S.A. Engineering College, Chennai, India

³Assistant Professor, Department of IT, S.A. Engineering College, Chennai, India

ABSTRACT: The successful implementation of customer relationship management (CRM) is not easy and seems to be a complex task. Almost about 70% of all CRM implementation projects fail to achieve their expected objectives. In today's era of competitive business and survival, there are some vital factors which lead to the success or failure of any business or organization.

A CRM analysis will also be performed in this project. The project depicts a dataset extracted from UCI (UC Irvine) repository preceded by **data pre processing, data cleansing, feature extraction**, splitting the data in the dataset into two parts - **training and testing data**, predicting results using a **machine learning algorithm**.

There are several machine learning algorithms such as **KNN (K-Nearest Neighbors), Linear Regression, Naive Bayes Theorem, Random Forest** and so on. Recruiting the right person for the right job is a crucial factor as it indirectly helps in providing a stable economy and give huge revenues or profits for the particular business organisation.

The project mainly emphasizes and depicts the key point that the organization's economic stability and **employee turnover** depends on the overall work productivity from an individual employee's perspective.

Keywords: CRM, UCI repository, data preprocessing, data cleansing, feature extraction, training and testing data, machine learning algorithm, KNN, Linear Regression, Naive Bayes Theorem, Random Forest, employee turnover.

INTRODUCTION:

The main perspective for this project is understanding and implementing the CRM (Customer Relationship Management) Analysis but in a different way that is, by analysing the data attributes in the UCI dataset especially during the pandemic time as a "work from home" culture has been evolved unexpectedly leading to greater difficulty in maintaining customer relationships due to various reasons. The existing dataset does not analyse the attributes in terms of the pandemic which has been identified as a motive to implement the proposed one.

Why CRM analysis is crucial for an organisation? Yes, mainly for effective analytical processes. The main function of analytical CRM is to analyze customer data so that management can better understand market trends and customer's wants and needs. The goal of analytical CRM is to improve customer satisfaction.

CRM analytics gives you insights about your customers and how well your sales and customer service teams are reaching them. CRM analytics helps you monitor your customer service efforts, validate your customer data, analyze your customers' habits and generate better leads.

Since the 1990s, customer relationship management (CRM) has emerged and grown rapidly in the business environment. In fact, CRM evolved from the Sales Force Automation (SFA) market, which was born of contact management.

Regardless of their type and size, businesses have been motivated to implement CRM to develop and manage a strong relationship with customers more effectively in an attempt to gain a comparative advantage. An enhanced relationship quality with customers can ultimately lead to increased customer satisfaction and an organization's profitability. In addition, the fast growth of the internet and its advanced technologies has greatly increased the opportunities for marketing and has transformed the way relationships between organizations and customers are managed.

In short, CRM technological initiatives are commonly implemented in functional departments, customer services and support, sales and marketing to maximize organization's profit. Hence, the organizations invest heavily in the CRM's projects for gaining a better understanding of the customers and for responding quickly to their requests and needs.

PROJECT OBJECTIVE:

The main advantage behind the project was to identify specified dataset attributes and to find the employee work ratio and other employee attributes considering the current pandemic. In this project, one can also demonstrate the salary hike and employee attrition ratio and employee performance in terms of individual company and business unit analysis.

The outcome of this project would enable the audience to grasp and implement certain forms and tricks and techniques of dataset analysis and maintenance along with an accurate outcome. The main objective is to implement CRM features pertaining to the pandemic scenario in addition to the analysis of the derived UCI dataset and the new dataset is termed as proposed dataset.

CRM and customer retention have increasingly been recognized as important managerial functions especially in saturated and increasingly competitive sectors. Every business firm knows that it costs far less to hold on to existing customers than to acquire new ones.

There is limited research available that investigates the effect of customer knowledge and orientation. In this backdrop the study tries to determine the impact of employee perception towards CRM practices (customer knowledge and customer orientation) on customer retention. Customer knowledge management is one of the important processes which utilize various knowledge management concepts and technologies in capturing the customer knowledge for developing better business relationships.

PROJECT SCOPE AND MOTIVATION:

The absence of additional required information pertaining to the present pandemic indicates the motivation behind executing this project in our employee dataset. The scope pertains to an accurate and a well-balanced analysis and interpretation of relationships, entities, attributes and entity sets within the given dataset.

The managers should give more emphasis on promoting customer-oriented values and behaviours among employees that will encourage in developing customer-oriented culture. Customer experience and satisfaction in the current business world are crucial to retain and attract new customers. However, customers get confused by the unlimited sources of marketing and information through the internet.

As such, organizations require an all-inclusive strategy to offer their current and potential customers channels for direct communications and a good experience. Hence, organizations must remain consistent in how they focus on selling processes, create new products and services, and even communicate company messages, among other factors.

Information technology has been proven to give organizations various advantages within the customer relationship management area. Organizations can enhance their customer experience through a customer relationship management system by fostering and enhancing customer loyalty and satisfaction. Additionally, a customer relationship management system also provides the organization with a competitive edge while increasing revenue and sales.

Machine learning, at its core, is an artificial intelligence category that aims at extracting insights or knowledge from a series of data or observations. It analyzes consolidated data and provides the business user with analytics and insights based on this data while also adapting and learning from every data set. Machine learning utilizes an algorithm or technique that requires minimal time to analyze all available data with greater accuracy and feature analysis.

The main CRM objective is augmenting and enhancing customer relationships in ways that an organization is able to achieve revenue growth, customer loyalty, as well as customer lifetime value through effective customer retention efforts. There is a need to highlight the integration between the different functions or dimensions when defining customer relationship management since it separates the phrase relationship marketing from customer relationship management. Relationship management can be defined as the overall customer relationship management aiming at long-term shareholders' value.

LITERATURE REVIEW:

1.) "Brave New World? On AI and the Management of Customer Relationships"-Barak Libai, Yakov Bart, Sonja Gensler, Charles F.Hofacker, Andreas Kaplan, Kim Kötterheinrich, Eike Benjamin Kroll(Year:2020).

The review is collected and analyzed from various journals pertaining to the topics such as "Employee Attrition Rate Prediction Using Machine Learning Approach", "Customer Relationship Management Approach Using Machine

Learning(CRM)".The machine learning algorithms in the existing system used are linear regression or Naive Bayes algorithm.

The proposed system uses the classification machine learning algorithm. The existing system uses either Naive Bayes Theorem or K-Nearest Neighbours(KNN) algorithms as referred from the journal "Employee Attrition Rate Using Machine Learning Approach.

The existing system also follows a traditional approach to Customer Relationship Management(CRM) System because it consists of a limited number of features in the dataset.

On the other hand, the proposed system using additional features in the dataset has two benefits- it improves CRM within an organization and it improves employee performance analysis. This project uses a decision-tree based classification algorithm known as "Random Forest" algorithm.

2.)"A systematic review for the determination and classification of the CRM critical success factors supporting with their metrics"-Marwa Salah Farhan, Amira Hassan Abed, Mahmoud Abd Ellatif (Year:2018).

It is estimated that 91% of companies with more employees use a Customer Relationship Management (CRM) system, compared with 50% of companies employing fewer people according to a journal case study analysis.

While these tools have become an integral part of the sales processes of businesses, the advances in AI technology impact this solution.AI technology is necessarily a part of CRM Systems. As CRM tools become more intelligent, they offer more accurate sales insights and help companies make better decisions in sales processes.

The CRM metrics reveal that the on-demand customer services and facilities are unavailable at most times. Feature extraction involving these attributes are missing in the dataset which makes it quite difficult to analyze, accumulate and understand various customer dissatisfaction factors.

This results in a kind of statistical analysis that fails to bridge the gap between employee performance metrics and customer satisfaction metrics. Exploring these areas with the help of traditional algorithms such as linear regression for instance, might lead to less accurate results.

Usage of better algorithms such as classification algorithms like Random Forest, for instance in future experiments for the same would lead to more accurate results and statistical analysis process.

3.)"Determining the Role of Employee's Perception towards CRM Practices and Customer Retention"" -Suhali Bhat,Mushtaq Ahmed Darzi(Year:2016).

A Customer Relationship Management(CRM) system allows businesses to monitor and analyze their relationships with customers, suppliers and employees. It aims to increase sales efficiency and profitability by improving and maintaining these business relationships.

By collecting existing and potential customer data, CRM tools can access customer's interaction histories and sales data. This enables CRM systems to identify the required services and products to improve sales processes.

The employee's performance might sometimes not continue to contribute to the customer's needs as an employee's perspective focuses on the aspect that whether the "Project is built right?" whereas from a customer's perspective it depicts whether the "Right Project is built?".

Hence, the employee only does the process of verification whereas the customer performs the validation task.

CRM systems reveal according to the recent estimates that the employee's relationship with customers using the CRM(Customer Relationship Management) software is going on a steady rate in a company's statistical analysis performed in the last quarter after performance appraisal being conducted in a US-based company in Mexico. It has hence, been depicted that atmost 50% of the employees do project verification but not validation and it has led to bad relations with customers hence losing many.

4.)"Employee satisfaction, human resource management practices and competitive advantage(The case of Northern Cyprus)"-Hamzah Elrehail, Ibrahim Harazneh, Mohammed Abhujeeleh, Amro Alzghoul, Shaker Alnajdawi(Year:2019).

Develop more effective marketing strategies. Discover new sales opportunities. Provide more effective support services in terms of CRM. One more important factor is employee attrition rate in terms of company.

Amongst the significant issues that corporate leaders have to deal with within an organization is the decline in proficient employees. Certain human resource practices have led to a major influence and often end up in work balance and relationship balance metrics.

This statistical feature extraction analysis technique reveals that-it has learnt to balance in terms of verification and validation processes and the two processes must go continuously hand-in-hand. A statistical analysis conducted as an experiment while analysis of these attributes and deriving an Entity Relationship Diagram(ER Diagram) for the same revealed a positive result.

It considered the relationship attributes and derived whether the two or more attributes within the dataset obtained are directly or indirectly proportional to each other.

5.)"Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches" -Raj Chakraborty, Krishna Mridha, Rabindra Nath Shaw, Ankush Ghosh(Year:2019).

The decline is primarily attributed to extreme work pressure, dissatisfaction at work and ignored mental health issues such as depression, anxiety, and so on. This is known as Employee Attrition Rate or Employee Churn Rate.

For instance, given the amount of stress employed people go through, focus on the state of mind has gained much-needed transaction. This model aims to predict the overall statistical employee performance analysis and feedback in addition to analysing attributes such as Work from Home performance, etc especially applicable to the current pandemic scenario.

Thus, employee turnover rate has had a negative impact due to lack of proper analysis of dataset attributes for the entity set termed as "employee" according to this article.

A case study suggests that the company which induces a positive attrition rate faces fewer negative challenges.

On the other hand, companies which induce and follow the principles of meagre or negative attrition rate might need greater challenges. For example, the meagre employee attrition rate principle induces a drastical side-effect on the workforce and workplace productivity.

This has a direct and significant relationship with the attrition rate within the organization.

6.)"An Approach for Predicting Employee Churn by Using Data Mining" -İbrahim Onuralp Yiğit,Hamed Shourabizadeh(Year:2017).

This model aims at predicting the employee attrition rate and the employees emotional assessment in an organization. With the help of various machine learning algorithms this analysis is carried forward.

The algorithms include K-Nearest Neighbours(KNN),Linear Regression, Support Vector Machine(SVM),Naive Bayes theorem, classification algorithms. Classification algorithms include random forest for example. Another popular example is that of the Decision tree Classifier algorithm.

Data mining techniques such as-

- Classification analysis
- Association rule learning
- Anomaly or outlier detection
- Clustering Analysis
- Regression Analysis

Listed above can also be used to gain an in-depth and significant relationship, derived relationships along with entities and entity relationships among various dataset attributes.

Extract-Transform-Load(ETL) method can also be used for the same. Various ETL tools are available.

It simply performs the key processing. An ETL tool extracts the data from different RDBMS source systems, transforms the data like applying calculations, concatenate, etc and then loads the data to the Data Warehouse System.

7.)"Customer relationship management:digital transformation and sustainable business model innovation" - Hermenegildo Gil-Gomez, Vicente Guerola-Navarro(Year:2020).

CRM(Customer Relationship Management) has a mostly direct or sometimes an indirect relationship with the Employee Turnover dataset attributes. Employee turnover can be described as a constant decline in the workforce due to retirement, death, or registration.

Meagre attrition results in the stagnation of ideas in the workplace.Positive attrition is considered beneficial as it generally results in incapable and less productive employees quitting the organisation.

This is an apt example for employee turnover within an organization case study.Meagre attrition rate also does not promote intellectual growth caused by exposure to new fresh recruits new ideas.

8.)Customer Relationship Management: Advances, Dark Sides, Exploitation and Unfairness" -Adam Rapp,Jessica Ogilvie(Year:2015).

A negative higher attrition rate may result in less productive workforce leading to serious consequences for the organization. On the other hand, there is another alternative way to reduce it.

Guess what it is? By training and identifying weaker areas and skills of the employees and training them to become more efficient. It has a two-factor advantage. One is that employee attrition rate, except for unforeseen circumstances, decreases rapidly. The other one is that a greater productive workforce is being built.

Sometimes, the analysis leads to a negative side-effect with respect to CRM(Customer Relationship Management).

A proper maintenance and data preprocessing is essential for an effective CRM software and if not maintained properly it can lead to a disastrous imbalance of analysis and relationship extraction.

9.)"Customer Relationship Management and Firm Performance"-Ali Ahani,Nor Zariah Abdul Rahim,Mehrbakhsh Nilashi(Year:2017).

The employees have a tough time dealing with their job dissatisfaction which develops insecurity and frustration for the job over a period of time indirectly. One of the case studies shows how to reduce employee attrition rate.

A US-based company reduced it's employee churn over rate drastically in a period of 2 years through skill training and counselling for all it's employees who were likely to fall in trap. and that was a very motivating and a real life example of a positive attrition rate.

To improve CRM and reduce employee attrition rate at the same time the organization can follow any one of the below principles-

- One way is to improve the incompetent employees is to provide adequate skill training based on the necessary prerequisites and skillset required in order to perform better and yield benefits for the organization.
- Another way is to follow the principle of only positive attrition rate after a deeper data mining or any other data analysis techniques in order to have lesser negative side-effects.
- Data Mining requires an extraction of required preprocessed dataset, transforming the dataset into a machine learning algorithm in order for the machine to analyze and process it.
- Followed by loading the data to be analyzed thus yielding accurate results.

10.)"Survey:Employee Attrition Rate Prediction Using Machine Learning"-Rohit Karale,Vishal Kolekar,Swati Sonawane,Aniket Karate,Balaji Bodkekar(Year:2020).

The employees had a limited bucket list during this pandemic. A case study demonstrates the absence of adequate resources to provide work productivity within the organisation especially during the pandemic period.

Limitations in case of a Work from Home scenario include poor Internet Connectivity and Bandwidth for an instance. Hence, the attributes to be added to the derived dataset must include the attributes related to the pandemic scenario in order for better data analysis.

Higher employee attrition rate might end up losing competent employees thus losing customer relationships and clients across the globe.

A case study in this article suggests that on the basis of a CRM (Customer Relationship Management) software at a company, the attrition rate was 5.1% in the F1 sales quarter.

The later analysis revealed that the company's database (CRM Software) in this case did not maintain and derive the latest updates regarding dataset attributes for employee entity dataset.

This resulted in an adverse and a negative side-effect thus affecting customer relationships across the globe and the company lost a lot of investing business clients during that year.

Hence, this reveals that a proper survey, maintenance and data exploratory analysis needs to be done for a less adverse effect.

EXISTING SYSTEM:

The existing system consists of a set of features pertaining to the scenario where an employee of the organization does the work in a regular corporate environment. This dataset does not have a certain set of features for a "work from home" scenario in terms of the pandemic. No CRM analysis attributes also included.

Various realistic issues occurring during this phase can be explained. As we all know, that the work environment for every employee has undergone a drastic change after the emergence of the COVID-19 pandemic since the year 2019.

PROPOSED SYSTEM:

The derived system consists of a set of employee performance attributes using the "regular office scenario" such as employee personal details, employee sick leaves applied, employee appraisal, employee CRM rating, employee salary, employee age and experience, employee actual and predicted working number of hours from office, and so on.

The proposed system consists of a set of the missing features so that the organization can work in a productive manner even during the course of the pandemic using the "work from home" set of features added in the proposed dataset in addition to the features of the existing dataset. Inclusion of features for employee's work from home performance and CRM analysis especially, with respect to the current scenario in real-time.

In this project, a classification algorithm known as the "Random Forest" algorithm is used. It is a supervised machine learning algorithm that is widely used in Classification and the Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

Classification algorithms such as random forest, decision-tree classifier are used in the proposed system so that it yields a higher and an accurate prediction compared to that of traditional algorithms like Support Vector Machine(SVM) algorithm, for example.

In order to overcome this problem, additional features are also included in the proposed dataset apart from the features in the existing dataset which makes it easier for feature analysis and extraction especially, from an organization's perspective.

In this project, additional steps are included apart from the problem statement identification and analysis. So, it yields two benefits for the organization-increases work productivity,and follows all COVID-19 protocols.

Additional attributes include a set of derived features for example, like covid vaccination dosage taken or not in order to comply with the covid protocol appropriate norms for example. This would enable the problem to be easily identified and resolved.

Understanding and comparing the two scenarios for an employee in terms of his or her work productivity and overall employee turnover would also enable them to analyze business profits, organization productivity and a smooth cooperation within the management.

It would further enable the organization to take a decision regarding whether the employee must be retained or not, performance appraisals which in turn helps them to follow COVID-19 appropriate norms such as social distancing during this pandemic time.

Organization can yield a lower employee attrition rate, higher work productivity and a better CRM(Customer Relationship Management) system by using the proposed system.

SYSTEM REQUIREMENT:

HARDWARE REQUIREMENTS-

- Intel i3 core processor.
- Computer system.

SOFTWARE REQUIREMENTS-

- Google Colaboratory.
- UCI repository.
- Python Libraries.
- Google Drive.

SYSTEM ARCHITECTURE:

The system architecture is represented by means of an ER Diagram Model. Here, "ER" refers to "Entity Relationship model". It has entity set, relationship and dataset attributes or dataset entities of the UCI Dataset.

Entity sets and Entities are represented as rectangle boxes in standard notation. Entities might be dependent ,independent, strong or weak entity. Each entity consists of a set of attributes which are it's characteristics. They are connected to each other using symbolic "arrow" notation.

The following are mentioned below in reference to the UCI dataset taken to implement this project.

- **Entity set-** Employee, Company.
- **Entity Relationship-** Here, the employee works for the company. Hence, "works for" is the final relationship.
- **Entities-**Age, Attrition, Business Travel, Daily Rate, Department, Distance From Home,

Education, Employee Count, Environment Satisfaction, Gender, Hourly Rate, Job

Involvement, Job Level, Job Role, Job Satisfaction, Marital Status, Monthly Rate, Num

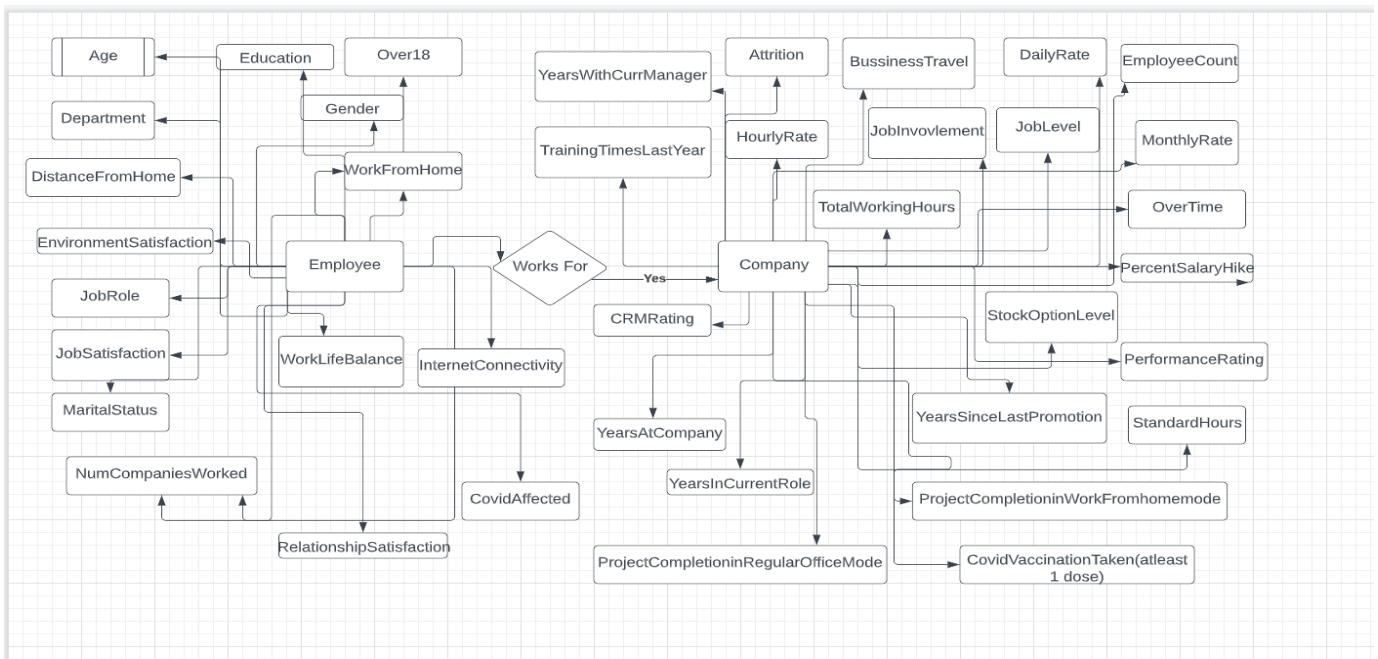
CompaniesWorked,Over18,OverTime,PercentSalaryHike,PerformanceRating,Relatio-
nshipSatisfaction,StandardHours,StockOptionLevel,TotalWorkingHours,TrainingTim-
esLastYear,WorkLifeBalance,YearsAtCompany,YearsInCurrentRole,YearsSinceLast

Promotion, YearsWithCurrManager, CRM Rating, Work From Home, Internet

Connectivity,ProjectCompletioninWorkfromhomeMode,ProjectCompletioninRegularOfficeMode,Covid Affected,
Covid Vaccination Taken(atleast 1 dose).

ER DIAGRAMS-

1.EXISTING SYSTEM-

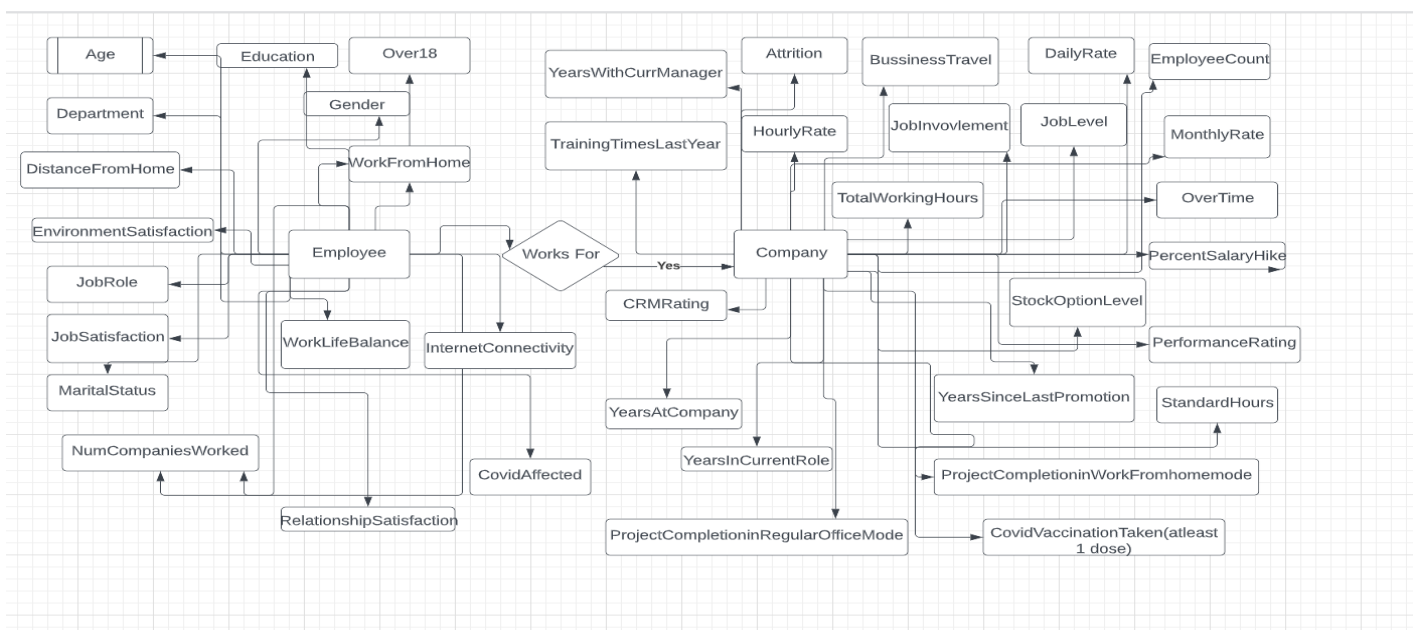


2.PROPOSED SYSTEM-

A significant phenomena of machine learning in the proposed system-

Several sources have tried defining machine learning. Machine learning is the science of programming computers to act and learn similar to humans and automatically enhance their learning by inputting information and data in real-world interactions and observations.

Machine learning can also be defined as a well-recognized and mature computer science research area that focuses mainly on discovering patterns, models, and other regularities within datasets. It can also be broadly defined as computation models or methods that rely on data or experience to make accurate forecasts or improve performance.



Real CRM began early in this century. CRM improved in recognizing the customers' needs and behaviors .CRM developed new strategies that accommodated work between understanding, sharing information and going to increase customer satisfaction.

Advanced solutions were flexible for different industries and they were feasible to use data and information in a dynamic way. Flexibility of the Internet makes easy relationships between sales, marketing and customer service.

CRM can get more advantages from high technology corporations, financial services and telecommunication industries. Business services can track the level of customer satisfaction, changing habits on purchasing products and make a product feedback to the organizations. One of the companies that released Global CRM was Oracle in 2002.

Business analytics and intelligence models can predict customers with a high likelihood to churn based on analysis of customer's behavioral, demographic, and personal data to provide customer-oriented and personalized marketing campaigns to achieve high satisfaction rates. The customer relationship lifecycle in business has four main stages, including customer acquisition, customer attraction, customer retention, and even customer development.

EXPERIMENTAL RESULTS:

Dataset-

1.Existing Dataset-

	B	C	D	E	F	G	H	I	J	K	L	M
1	BusinessTravel	DailyRate	Department	DistanceFromHome	WorkEducation and Implant	EducationalField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
2	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female		94
3	Travel_Frequent	279	Research & Development	8	1	Life Sciences	1	2	3	Male		61
4	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4	Male		92
5	Travel_Frequent	1392	Research & Development	3	4	Life Sciences	1	5	4	Female		56
6	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1	Male		40
7	Travel_Frequent	1005	Research & Development	2	2	Life Sciences	1	8	4	Male		79
8	Travel_Rarely	1324	Research & Development	3	3	Medical	1	10	3	Female		81
9	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	1	11	4	Male		67
10	Travel_Frequent	216	Research & Development	23	3	Life Sciences	1	12	4	Male		44
11	Travel_Rarely	1299	Research & Development	27	3	Medical	1	13	3	Male		94
12	Travel_Rarely	809	Research & Development	16	3	Medical	1	14	1	Male		84
13	Travel_Rarely	153	Research & Development	15	2	Life Sciences	1	15	4	Female		49
14	Travel_Rarely	670	Research & Development	26	1	Life Sciences	1	16	1	Male		31
15	Travel_Rarely	1346	Research & Development	19	2	Medical	1	18	2	Male		93
16	Travel_Rarely	103	Research & Development	24	3	Life Sciences	1	19	3	Male		50
17	Travel_Rarely	1389	Research & Development	21	4	Life Sciences	1	20	2	Female		51
18	Travel_Rarely	334	Research & Development	5	2	Life Sciences	1	21	1	Male		80
19	Non-Travel	1123	Research & Development	16	2	Medical	1	22	4	Male		96
20	Travel_Rarely	1219	Sales	2	4	Life Sciences	1	23	1	Female		78
21	Travel_Rarely	371	Research & Development	2	3	Life Sciences	1	24	4	Male		45
22	Non-Travel	673	Research & Development	11	2	Other	1	26	1	Female		96
23	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	3	Male		82
24	Travel_Rarely	419	Research & Development	7	4	Life Sciences	1	28	1	Female		53
25	Travel_Rarely	391	Research & Development	15	2	Life Sciences	1	30	3	Male		96
26	Travel_Rarely	699	Research & Development	6	1	Medical	1	31	2	Male		83
27	Travel_Rarely	1324	Research & Development	3	2	Other	1	32	3	Female		58

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked	Over18	OverTime	PercentSalaryHike	PerformanceRating	CRM Rating	StandardDeviation
2	3	2	Sales Executive	4	Single	5993	19479	8	Y	Yes	11	3	1	
3	2	2	Research Scientist	2	Married	5130	24907	1	Y	No	23	4	4	
4	2	1	Laboratory Tech	3	Single	2090	2396	6	Y	Yes	15	3	2	
5	3	1	Research Scientist	3	Married	2909	23159	1	Y	Yes	11	3	3	
6	3	1	Laboratory Tech	2	Married	3468	16632	9	Y	No	12	3	4	
7	3	1	Laboratory Tech	4	Single	3068	11864	0	Y	No	13	3	3	
8	4	1	Laboratory Tech	1	Married	2670	9964	4	Y	Yes	20	4	1	
9	3	1	Laboratory Tech	3	Divorced	2693	13335	1	Y	No	22	4	2	
10	2	3	Manufacturing D	3	Single	9526	8787	0	Y	No	21	4	2	
11	3	2	Healthcare Repr	3	Married	5237	16577	6	Y	No	13	3	2	
12	4	1	Laboratory Tech	2	Married	2426	16479	0	Y	No	13	3	3	
13	2	2	Laboratory Tech	3	Single	4193	12682	0	Y	Yes	12	3	4	
14	3	1	Research Scientist	3	Divorced	2911	15170	1	Y	No	17	3	4	
15	3	1	Laboratory Tech	4	Divorced	2661	8758	0	Y	No	11	3	3	
16	2	1	Laboratory Tech	3	Single	2028	12947	5	Y	Yes	14	3	2	
17	4	3	Manufacturing D	1	Divorced	9980	10195	1	Y	No	11	3	3	
18	4	1	Research Scientist	2	Divorced	3298	15053	0	Y	Yes	12	3	4	
19	4	1	Laboratory Tech	4	Divorced	2935	7324	1	Y	Yes	13	3	2	
20	2	4	Manager	4	Married	15427	22021	2	Y	No	16	3	3	
21	3	1	Research Scientist	4	Single	3944	4306	5	Y	Yes	11	3	3	
22	4	2	Manufacturing D	3	Divorced	4011	8232	0	Y	No	18	3	4	
23	2	1	Sales Represent	1	Single	3407	6986	7	Y	No	23	4	2	
24	3	3	Research Direct	2	Single	11994	21293	0	Y	No	11	3	3	
25	3	1	Research Scientist	4	Single	1232	19281	1	Y	No	14	3	4	
26	3	1	Research Scientist	1	Single	2960	17102	2	Y	No	11	3	3	
27	2	5	Manager	2	Divorced	4004	10725	4	Y	No	44	3	4	

	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AI
1	StandardHours	EmployeeRetainingLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager	WorkFomHomeCapacity	InternetConn
2	80	0	8	0	1	6	4	0	5	4	
3	80	1	10	3	3	10	7	1	7	2	
4	80	0	7	3	3	0	0	0	0	3	
5	80	0	8	3	3	8	7	3	0	3	
6	80	1	6	3	3	2	2	2	2	2	
7	80	0	8	2	2	7	7	3	6	4	
8	80	3	12	3	2	1	0	0	0	1	
9	80	1	1	2	3	1	0	0	0	2	
10	80	0	10	2	3	9	7	1	8	5	
11	80	2	17	3	2	7	7	7	7	2	
12	80	1	6	5	3	5	4	0	3	3	
13	80	0	10	3	3	9	5	0	8	3	
14	80	1	5	1	2	5	2	4	3	3	
15	80	1	3	2	3	2	2	1	2	4	
16	80	0	6	4	3	4	2	0	3	3	
17	80	1	10	1	3	10	9	8	8	1	
18	80	2	7	5	2	6	2	0	5	2	
19	80	2	1	2	2	1	0	0	0	4	
20	80	0	31	3	3	25	8	3	7	4	
21	80	0	6	3	3	3	2	1	2	4	
22	80	1	5	5	2	4	2	1	3	3	
23	80	0	10	4	3	5	3	0	3	1	
24	80	0	13	4	3	12	6	2	11	2	
25	80	0	0	6	3	0	0	0	0	4	
26	80	0	8	2	3	4	2	1	3	1	
27	80	0	0	0	0	0	0	0	0	0	

2. Proposed Dataset-

	AI	AJ	AK	AL	AM	AN
1	WorkFomHomeCapacity	InternetConnectivityIssues	Project completion in Work from home Mode	Project completion in Regular Office Mode	Covid Affected	Covid Vaccination Taken(atleast 1 dose)
2	4	0	5	4	NO	YES
3	2	3	4	2	YES	NO
4	3	4	3	3	YES	NO
5	3	3	3	3	NO	YES
6	2	2	2	2	NO	NO
7	4	4	4	4	YES	NO
8	1	1	1	1	NO	NO
9	2	3	0	3	NO	YES
10	5	3	1	3	YES	YES
11	2	3	2	3	YES	NO
12	3	2	4	2	YES	NO
13	3	3	5	3	NO	YES
14	3	3	2	3	NO	NO
15	4	4	3	4	NO	NO
16	3	3	4	3	YES	NO
17	1	1	5	1	NO	YES
18	2	2	0	2	YES	NO
19	4	4	5	4	YES	YES
20	4	4	1	4	NO	YES
21	4	4	0	4	NO	NO
22	3	3	0	3	YES	NO
23	1	1	2	1	NO	YES
24	2	2	2	2	NO	NO
25	4	4	4	4	YES	NO
26	1	1	1	3	YES	YES
27	0	0	0	0	YES	YES

This proposed dataset contains additional features known as data attributes or entities in addition to the existing attributes in the derived or existing dataset for better and an improvised dataset analysis pertaining to the real-world scenario.

CONCLUSION:

This project includes and emphasizes upon additional factors pertaining to the real world scenario of the COVID-19 pandemic. Machine learning-based data mining techniques are a combination of machine learning, artificial intelligence, and classical statistics. Advancement in the artificial intelligence field has led to the development of machine learning techniques that have proved useful when automating tedious but crucial discovering patterns activities within databases.

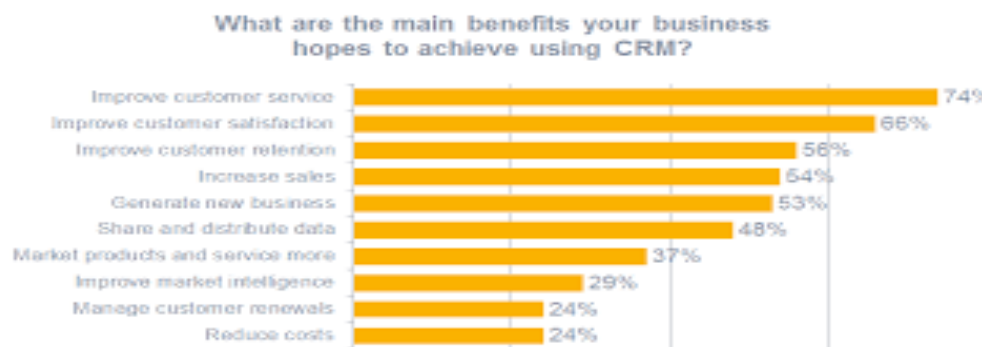
In the past, for a general example let us consider the scenario where business analysts used general statistical techniques and Online Analytical Processing analysis to identify a pattern within organization data. However, this method had some challenges. For Online Analytical Processing, it required a user to have query objectives' subjective definition. In comparison, statistical techniques required the dataset to adhere to inflexible distribution standards, allowing statistical modeling methods.

The dataset has been derived from UCI repository and companies can use the proposed dataset in order to improve CRM and employee performance analysis as the proposed dataset would be made available in Kaggle or any other public dataset repository after the project is executed. It is an IBM HR Analytics dataset. Additional features for the “work from home” scenario were included to make the proposed dataset. This would enable the organization to make efficient decisions during the course of the pandemic.

FUTURE ENHANCEMENT:

This project has been implemented using a traditional classification algorithm(Random Forest Algorithm).It is suggested to experiment the proposed dataset with customized machine learning algorithms.This dataset can be analyzed using deep learning or big data analytics approaches in future.

It is also suggested further to add some more features in the dataset in order to obtain additional information to create a dynamic database approach with certain customized approaches.Further data exploration and data mining approaches can be applied and experimented in the proposed system or proposed dataset.



REFERENCES:

- 1.)”Brave New World? On AI and the Management of Customer Relationships”-Barak Libai,Yakov Bart,Sonja Gensler,Charles F.Hofacker,Andreas Kaplan,Kim Kötterheinrich,Eike Benjamin Kroll(Year:2020).
- 2.)”A systematic review for the determination and classification of the CRM critical success factors supporting with their metrics”-Marwa Salah Farhan,Amira Hassan Abed,Mahmoud Abd Ellatif (Year:2018).
- 3.)”Determining the Role of Employee’s Perception towards CRM Practices and Customer Retention” -Suhali Bhat,Mushtaq Ahmed Darzi(Year:2016).
- 4.)”Employee satisfaction, human resource management practices and competitive advantage(The case of Northern Cyprus)”-Hamzah Elrehail,Ibrahim Harazneh,Mohammed Abhujeeleh,Amro Alzghoul,Shaker Alnajdawi(Year:2019).
- 5.)”Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches” -Raj Chakraborty,Krishna Mridha,Rabindra Nath Shaw,Ankush Ghosh(Year:2019).
- 6.)”An Approach for Predicting Employee Churn by Using Data Mining” -Ibrahim Onuralp Yigit,Hamed Shourabizadeh(Year:2017).
- 7.)”Customer relationship management:digital transformation and sustainable business model innovation” -Hermenegildo Gil-Gomez,Vicente Guerola-Navarro(Year:2020).
- 8.)Customer Relationship Management: Advances, Dark Sides, Exploitation and Unfairness” -Adam Rapp,Jessica Ogilvie(Year:2015).
- 9.)”Customer Relationship Management and Firm Performance”-Ali Ahani,Nor Zariah Abdul Rahim,Mehrbakhsh Nilashi(Year:2017).
- 10.)”Survey: Employee Attrition Rate Prediction Using Machine Learning”-Rohit Karale,Vishal Kolekar,Swati Sonawane,Aniket Karate,Balaji Bodkekar(Year:2020).