International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# EMPLOYEE PAYROLL DATA ANALYSIS USING MACHINE LEARNING

**[1]MOHAMAD KAIF M M, [2]ROOPA R**
[1] *Student*, Department of *MCA*, *BIET*, Davangere
[2] *Assistant Professor* Department of *MCA*, *BIET*, Davangere

**ABSTRACT:**
Every company is using machine learning to predict their employees' working styles and competence by estimating how long it will take them to complete a task. The effectiveness of various classification algorithms for predicting employee salary classes has been the subject of numerous recent studies. Salary prediction appears to be a challenging endeavor from the perspective of machine learning because of the low sample size, relatively high dimensionality, and presence of noise. Deeper architectures are required to address this and locate more useful features. Also, more information investigation and information handling can be commonsense to make the forecast model go past the relationship and accuracy guidelines by highlight extraction methods.

## 1. INTRODUCTION:

Since the last few decades, employers and individuals have wanted to keep records of their employees. Initially, they kept the records manually, but as the number of employees increased, it became more difficult to keep track of them and easily access them. The management of an employee's financial data, including their salary, wages, benefits, and deductions, is referred to as "payroll management." If needed later, that record ought to be kept. The purpose of this software is to alleviate or, in some cases, lessen the difficulties faced by the existing system.

A payroll system is a method by which an organization pays its workers. It incorporates essential compensation and different offices that a worker gets from an association. Calculating salary, expenses, and holiday-leave pay with the payroll system makes it simple to generate accurate payslips.

The suggested solution was created to address issues with the manual system's use, including excessive administrative burden, handling overtime, incorrect employee classification, inefficient timekeeping, monitoring employee absences, and delayed payroll processing. The goal of this web application is to minimize mistakes made when inputting data.

## 2. LITERATURE REVIEW

Marcus Atish D Rozario (2018) implemented a web-based payroll system that was built with PHP, HTML, and JavaScript. The database was designed by PHP MyAdmin, and Windows 10 was used as the operating system. Computerized records can be maintained and viewed using this web application without having to deal with duplicate entries. This system is only used by admin, and it is used in small businesses with few employees.[1]

Arjun V. Singh et al. (2016) implemented a desktop-based payroll system with Microsoft Access 2007 SQL server 2008 as the backend and VB.net as the frontend. It was developed as a college management system, with admin access only.[2]

Kritika Mahajan et al. (2015) implemented a desktop payroll system that uses JSON and Ajax for data parsing and frontend development, C# for backend development, and HTML, CSS, and jQuery for frontend development.[3]

Poonam Deep Kaur et al.'s computer-based payroll system was developed with PHP, HTML, CSS, JavaScript, and MySQL, and the database was designed with MySQL. Ed. Access to login and logout is controlled by the administrator, and the head of HR is in charge of other employees' data.[4]

Singh et al. (A web-based payroll and leave system was proposed in 2017). APIs are used to connect the system to other systems. In such a foundation where it is carried out, new elements can be added which probably

won't have been there toward the start when the framework is being sent and it won't influence the honesty of the framework. It is also set up in a way that makes it work and connect with the biometric attendance system by itself.[5]

Rahman et al. (2017) suggested an automated system in their work. There are six distinct modules in the system. The employee management system is one of the modules; others include the employee payroll system, employee performance management system, employee leave management system, employee monitoring system, and employee attendance management system. Employees, administrators, and operators use the system. Depending on what is expected of them, each person has different access to the system.[6]

Tharanya and Senthil Raja (2020) introduced a web-based smart attendance and payroll system in which each employee is given a unique employee number and their information is stored in a database. Every day, an administrator logs into the system to add and remove employees. This step catches labourer's time and gives an everyday movement of time records as a report or for additional use, for example, the month-to-month wage computation by finance division programming participation the board.[7]

Barcodes were used by Kimani and Nyatuka (2018) to verify users' attendance. The payroll calculation uses the barcode that users scan on their pass ID to enter and exit. PHP, MySQL, HTML and notepad++ were utilized as instruments, and programming prototyping philosophy was utilized.[8]

Patra and others (2021) presented a framework that gathers the participation of laborers with a biometric scanner. The cloud is simultaneously uploaded with the collected attendance data. Employees and the organization alike can use their respective applications to access this data. With the help of the software used to calculate payroll, the company will be able to work with employee data by taking attendance and doing everything else.[9]

Prasad et al.'s research (2019), each employee receives a secure identification number that they use to authenticate. This number, in addition to other information like coordinates and images, is saved on their Android smartphones that also have the necessary APK files installed on them. For security reasons, the employees use their smartphones to log into the system using their credentials and take a picture. Representatives are expected to turn on their GPS on their Android cell phones as the cycle will not go in the event that it is switched off. Every representative's cell phone on getting to work interfaces with a passage WIFI and does participation and the records are utilized to ascertain their compensation.[10]

## 3. METHODOLOGY:

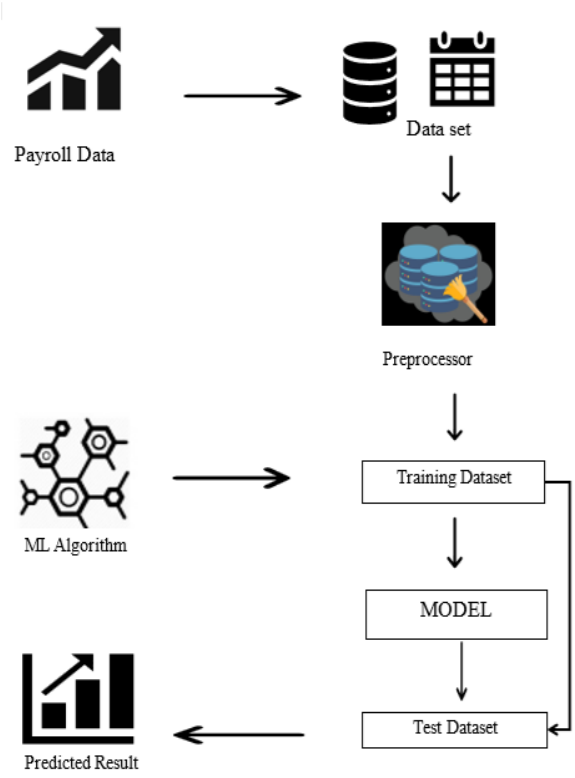



**Fig 1: System Architecture**

### 3.1 Data Collection

This paper by making use of the dataset provided by revolution analytics. The Train set, the Valid set, and the Test set each contain 60%, 20%, and 20% of the dataset.

### 3.2 Data preprocessing and Analysis

Prior to feature selection, data preprocessing is carried out in order to facilitate the effective implementation of the classification algorithm. Under International Diary of Designing Science and Progressed examining is performed to make the dataset adjusted to stay away from the biasing of the characterization calculation towards the greater part class. Information Pre-

Handling and Examination Stage Machine learning models significantly increment their exhibition through the essential period of information pre-handling.

### 3.3 Training Dataset

It simply predicts the training data's average salary for all employees. There is no bias toward any aspect. I tried to include the location, the company's name, and the job title in this model. I involved 'pandas' for my code. My file reads and computations were.

I used the job category for this model, which isn't very broad, as the data description and the company name show. I created a subset of data using the group by feature in Pandas, where in the average salaries for each category and company are represented in each row.

### 3.4 ML Technology's

The applied ML methods are examined in this section. K-Means, Irregular Timberland, Decision Trees, kth Nearest Neighbor (k-NN), and super Vector Machine (SVM)examples of relapse.

#### 3.4.1 Decision Trees

Decision trees are a powerful addition to an HR analyst's toolkit. They are supervised machine learning algorithms that can handle both categorical (classification) and numerical (regression) outcomes.

#### 3.4.2 K-Nearest Neighbor (KNN) Regression

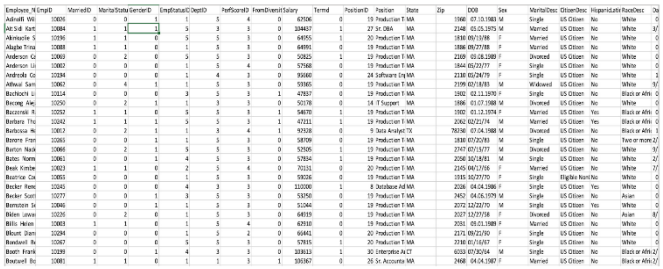
KNN is another algorithm that can be useful for payroll analysis.

It's a non-parametric method that predicts numerical outcomes based on the average of the 'k' nearest data points.

KNN works well when your dependent variable (e.g., salaries) is not normally distributed.

### 3.5 Dataset used

In the ML work process, text preprocessing is a significant stage. In essence, it entails cleaning the data so that only the relevant data can be extracted. The first we cleaning data is to look for duplicate or irrelevant data. We used a variety of hashtags to collect data over various time periods. It can oftentimes provoke overabundance tweets by something almost identical or different clients having different typical hashtags. We killed deceit in data arrangement stage itself.



**Tablet 1**: Data Set

The project's data flow is depicted in the above diagram. At first, the customer sends a request to the administrator to check the large amount of network data that needs to be checked.

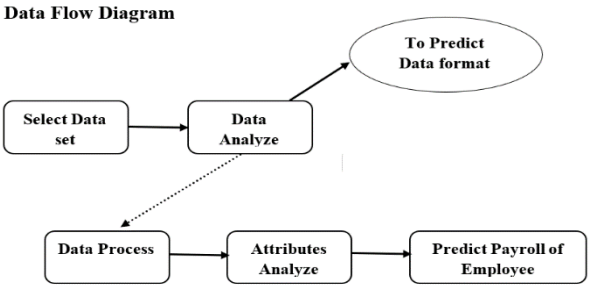



**Fig 2: Data Flow Diagram**

## 4. RESULTS AND DISCUSSION:

The process of gathering, organizing, and evaluating payroll-related data in order to derive useful insights is referred to as "payroll data analytics." The salary, bonuses, deductions, benefits, and tax information of employees are all included in this data. Payroll companies can gain a comprehensive understanding of their clients' payroll processes, identify patterns, and make informed decisions by utilizing advanced analytical methods.

**Algorithms work**

"Decision Tree Regression" was used to predict a person's salary.

A decision tree, a tool for making decisions, can be used to show all possible outcomes, such as outcomes, input costs, and benefits. Decision tree algorithms are included in a group of supervised learning algorithms. works with continuous output variables that are categorized.

A node's result is represented by a branch or edge, where a node has either:

1. Prerequisites (Decision Nodes)
2.Result (End Nodes)

Decision tree regression trains a model in the form of a tree to observe element properties and predict future dates in order to produce useful continuous results. The absence of a discrete output is indicated by a continuous output. H. An output that is more than just a discrete, recognizable number or set of values.

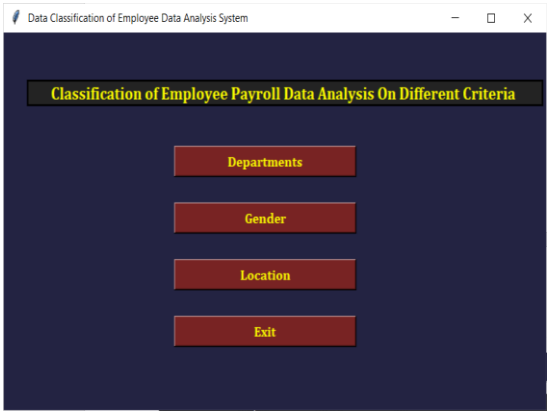



**Fig 3: Data Classification with Different Criteria's**


**Fig 4: Dataset Result Of Department Criteria's**

K-Nearest Neighbors (KNN)

The classification and regression algorithm KNN is both simple and efficient.

In our case, we'll use it for classification (attrition prediction).

Steps:

Compute distances between data points (employees) based on features.

For each employee, find the kth nearest neighbors (other employees) based on feature similarity.

Assign the majority class label (attrition or no attrition) among the neighbors to the target employee.

KNN works well when local patterns matter (e.g., similar employees tend to have similar attrition behavior).

Visualizing the KNN Process:

We won't create a full graph, but imagine connecting each employee to their kth nearest neighbors.

The decision boundary (attrition vs. no attrition) is influenced by the neighbors' labels.

Visualizing this can help us understand how KNN makes predictions.

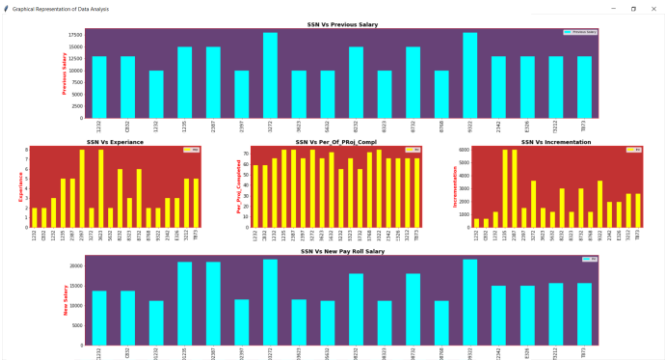



**Fig 5: Graphical Representation of Analyzed Employee Data**

## 5. CONCLUSION:

This undertaking is worked for various sorts of organizations and it is easy to use. One can effectively and sequentially manage all employee details by adapting this system. The database successfully stores all employee information, including login, registration, salary information, leave information, and tasks. Employee information is also protected, and even if they require it, they can obtain it quickly. It too assists with deleting all human mistakes happened during manual approach. It can quickly calculate all payroll management costs. This application can update salary records as well as all salary, employee attendance, overtime, and leave issues, among other issues. As a result, this application will assist in automating that organization's business data.

**REFERENCE:**

[1] Marcus Atish D Rozario (2018) "Worker Data set and Finance The board Framework" Maulana Abul Azad College of Innovation Kolkata, Division of PC Application.

[2] "Automated Payroll System (A-Pay)" by Arjun V. Singh, Siddesh V. Chaphekar, and Yogesh S. Sawant, International Journal of Modern Trends in Engineering, Vol. 03, February 2016, Issue 02

[3] "A Review of Computerized Payroll System," by Kritika Mahajan, Shipla Shukla, and Nitasha Soni, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2015.

[4] Poonamdeep Kaur, Dr. Dinesh Grover "PC Based Finance Framework Execution for E-Administration at Punjab" Farming College, Branch of Software engineering. The International Journal of Engineering Development and Research Vol. 5, Number 3, Dec. 2012.

[5] "Employee Payroll Management System," Prabu S., Akash Goyal, Anmol Agrawal, and Siddhant Nagelia, International Journal of Innovative Science and Research Technology, Volume 3, Issue 4, April 2018.

[6] Dhan Amma Jagli, Ramesh Solanki, Parth Chandaranam "Finance The board Framework as SaaS" Proceedings of Public Meeting on New Skylines in IT - NCNHIT2013.

[7] Bangladesh Open University's School of Science and Technology's "Payroll Management System."

[8] "Survey on Various Automated Payroll System" by Sampada A. Chavan, Neelam D. Gaikwad, Vinay V.sakpal, Javed Bilakhia, and Rupali Pashte, International Journal for Research & Development in Technology, Vol. 17, March 7, Issue 3.

[9] Employee payroll information system on corporate web-based consulting engineering services, Hikmah IN, Muqorobin M (2020). 1(2):27–30 Int J Comput Inf Syst IJCIS

[10] Rubin J, Chisnell D (2008) Wiley, Hoboken, 2nd ed. Handbook of usability testing: how to plan, develop, and run effective tests