Evaluating the Impact of remote work on Employee Productivity and Satisfaction using Machine Learning Approaches

MSc Research Project
MSc Artificial Intelligence for Business

Rajas Abhijit Zende
Student ID: x23352779

School of Computing
National College of Ireland

Supervisor: Dr Muslim Jameel Sayyed

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Rajas Abhijit Zende……. ……
Student ID:	X23352779……………………………………………………………………………………………………
Programme:	MSc in Artificial Intelligence for Business……………………………………………………………… **Year:** 2024-2025
Module:	MSC research project …….………
Supervisor:	Dr Muslim Jameel Sayyed ……………………………………………………………………………………………….………
Submission Due Date:	15/09/25……. ………
Project Title:	Evaluating the Impact of remote work on Employee Productivity and Satisfaction using Machine Learning Approaches …….………
Word Count:	9660 words ……………………………………… **Page Count** 26 pages

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	…Rajas Abhijit Zende ……
Date:	11/08/25 ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Evaluating the Impact of remote work on Employee Productivity and Satisfaction using Machine Learning Approaches

Rajas Abhijit Zende
Student ID x23352779

Abstract

Remote working model is being implemented by almost every organization after Covid 19 pandemic. The implementation of remote work model offers flexibility and a better work life balance to employees with some studies indicating that employees who work remote have high productivity and satisfaction scores. On the contrary some studies discuss the negative effects of working remote such as stress, lack of social contact which harms the mental health of employees further affecting the productivity and satisfaction of employees. In this study we investigated whether remote work has a direct impact on employee efficiency. By using a dataset consisting of 10000 employees records we conducted feature engineering and developed 2 machine learning models Random Forest model and LGBM model. Also, in this study we conducted feature engineering and developed complex features and a target variable by combing performance score and satisfaction score along with 2 remote work interaction terms that were successful in capturing the indirect impact of remote work on employee efficiency. The LGBM model that was trained and received an exceptional R2 score of 0.8099. in our study a key finding revealed that remote work didn't have a direct impact on the efficiency of employees but had an indirect impact through remote work interaction features which when combined contributed to (30.10%) towards the model's performance. This study will be useful for organizations who view the impact of remote work of productivity and satisfaction is similar for all the employees. Further enabling them to create personalized remote work policies.

1 Introduction

After the Covid 19 pandemic most of the organizations have adopted remote and hybrid working models changing the traditional ways of organizations in measuring employee productivity and satisfaction. Remote work model is being implemented by multiple organizations in different sectors such as Information technology, finance, and human resources. The adoption of remote work model offers flexibility and a better work life balance however it raises a question of its impact on the performance and satisfaction of employees. Some studies indicate the employees working in remote environments report high satisfaction and productivity while some studies report that employees working in remote environments often have high levels of stress, communication barriers and a lack of social networking which affects their productivity and satisfaction. Rehman et al. (2025) found that employees who were experienced were highly satisfied with remote work model. However, the implementation of remote work may not be beneficial for all for e.g. new employees or employees having a low work experience may struggle to adopt to the remote working model due to lack of support and mentorship. These findings indicate a critical problem in current research methods as most studies consider remote working model as a key factor where remote work affects the productivity and satisfaction equally. Whereas remote work model affects the productivity and satisfaction of every employee differently based on their characteristics such as their collaboration with their team, their career journey. With the advancement in machine learning techniques and models the factors affecting the performance and satisfaction metrics of employees can be identified and, we can identify that does remote work really affects the performance and satisfaction of employees.

In our research we attempt to find out the factors affecting the performance and satisfaction of employees and does remote work have an impact on performance and satisfaction of employees.

Research questions.

1) To what extent does remote work have an impact on productivity and satisfaction metrics.

2) Does remote work directly affect the productivity and satisfaction of employees

3) What are the key factors that actually affect the productivity and satisfaction of employees?

4) Can machine learning models help in finding out if remote work has an impact on employee productivity and satisfaction?

2 Related Work

In the related work chapter, we discuss the existing research conducted on remote employee productivity and employee productivity metrics. The literature review discusses 15-20 research papers and machine learning approaches used to assess the Impact of Remote work on employee productivity and performance metrics. In this literature review we discuss the strengths and weaknesses of each study and identify the gaps present in the studies which our research aims to address.

2.1 Assessing the impact of remote work using machine learning techniques.

Several studies have been conducted to determine how remote work impacts the productivity and wellbeing of employees. Rahman et al. (2025) conducted a comprehensive study to find out the impacts of remote work models on the productivity and mental health being of employees using machine learning models. In this study a dataset comprising of 5000 entries was used and included employees work related features and well being related features. Rahman et al. (2025) employed 10 machine learning models for finding the impact of remote work on employee’s productivity and wellbeing. In the study conducted logistic regression was the best performing model as it achieved an accuracy of 0.85 and found that employees who had high years of experience had high satisfaction and wellbeing in remote work. Rahman et al. (2025) concluded that experience of an employee is an important factor that helps the employee in adapting smoothly to remote working environment. However, Rahman et al. (2025) identified several limitations in the research and suggested the need for more complex features to for understanding the experience of employees in remote working model. Rahman et al. (2025) suggested that more work-related features need to be investigated that contribute to employee's wellbeing and productivity and by doing so the organizations can suitable remote working environment for its employees.

Jacob and Jayalakshmi (2025) analysed the impact of remote work on productivity and work life balance by using a baseline and federated learning models. Jacob and Jayalakshmi (2025) selected productivity and work balance factors and the frequency of breaks of employees. The federated machine learning model achieved an accuracy of 85% and an improvement of 9% over the baseline model implemented. Jacob and Jayalakshmi (2025) found that remote work can help in boosting productivity when employees are provided with flexible schedules and have a good work life balance.

Okushima (2022), conducted a study in Japan's provincial regions to determine the relationship between remote work mode and its factors and focused on partial remote work. Okushima (2022) conducted a web-based survey in Tokushima and other parts of Japan to understand the intention of remote work. Okushima (2022), gathered 400 survey responses and employed logistic regression model followed by decision tree model, a random forest model, Support Vector model and a neural network model for finding the intention of regular remote work and for finding the intention of partial remote work Okushima (2022), applied the same models. where in Regular remote work intention Random Forest and SVM emerged as the best performing models and for partial remote work intention Logistic regression and neural networks emerged as the best performing models. Okushima (2022), emphasized that regular remote work intention is dependent on total number of activities and partial remote work is dependent on the number of hours. Okushima (2022), concluded that more samples need to be collected to support the performance of the machine learning models and that it is necessary to tune the frequency of remote work.

Kumara et al. (2022) employed a machine learning approach to predict employee's choice of work setting after Covid 19. Kumara et al. (2022) developed and used a questionnaire to collect data from 250 employees working in different organizations to get their responses on their preferred working model. Kumara et al. (2022) conducted a comparative analysis with different machine learning models which included Random Forest, ANNs, Naïve bayes and an ensemble learning classifier. The ensemble learning classifier achieved the highest precision score of 1.00 followed by ANN and naïve bayes. Kumara et al. (2022) concluded that the size of the dataset needed to be increased and different bagging and boosting needed to be considered.

Zeeshan et al. (2025) talks about the applications of artificial intelligence into human resource management practices particularly in remote work. Zeeshan et al. (2025) discuss the benefits of using artificial intelligence techniques such as machine learning and predictive analytics to facilitate collaboration among teams in different working models. Zeeshan et al. (2025) also discusses the benefits of using artificial intelligence to enhance satisfaction and productivity among remote employees and with the help of artificial intelligence the employees working remotely can collaborate and interact with each other over a platform without miscommunication and the productivity of the employees can be increased if the regular routine tasks are automated with the help of artificial intelligence. Finally, Zeeshan et al (2025) discuss that the Artificial intelligence specialists and human resources teams must collaborate to implement artificial intelligence in organizations and how artificial intelligence tools have a vast potential to analyse the performance of employees and elevate their experience.

Seghyar et al (2024) discussed the impact of artificial intelligence tools on employees working in remote work settings. The impact of various artificial intelligence tools such as ASANA and Cortana on employee's performance and satisfaction was discussed. The implementation of artificial intelligence has led to the creating of various tools for employees working in remote environments. Different algorithms such as Asana, Trello are used for helping the employees in managing their tasks effectively. Seghyar et al.(2024) discussed about how artificial intelligence affects the satisfaction and performance of employees in remote work by automation of tasks and effective communication. In the study a financial firm had implemented AI tools after Covid 19 which included Asana and Cortana for the project management and automation of repetitive tasks. The main purpose of the study was to find out how well the remote employees have adapted to the artificial intelligence tools. Questionnaires were used to gather information. Employees using the artificial intelligence tools reported a 25% increase in productivity and 80% of the employees were able to adapt to the artificial intelligence tools with 20% employees facing problems and unable to adapt to the artificial intelligence tools. Seghyar et al (2024) concluded that there are benefits of implementing artificial intelligence tools in remote work model, but support and assistance must be provided to the employees who are facing difficulties in adapting to the artificial intelligence tools and the organizations must invest in providing training to all the employees working remotely.

Rehan et al. (2023) conducted a comparison analysis of working remotely vs working from office using machine learning algorithms to identify the work models preferred by software engineer employees. In this study a questionnaire was used to capture the work preference mode of the employees, and 174 samples were collected. classification algorithms were used which included Decision trees, random forests, Support vector machine and logistic regression. Decision tree and random forest classifier had the same accuracy score of 96.23% followed by support vector machine and logistic regression. In the study we conducted by Rehan et al. (2023) we observed that although the accuracy of the models was high, but the sample size was small which concerns the generalization of the results.

Gamage and Asanka (2022) used machine learning approaches to predict mental stress of IT employees in remote working environments. The study was conducted during the Covid 19 pandemic with remote IT employees. A questionnaire was used and responses from 481 employees were taken which included information on mental health and occupations. Different ML model such as logistic regression, catboost, RFT, SVM and catboost were used for classification. XGBoost achieved an accuracy score of 96.8% followed by catboost Naïve bayes and SVM achieved a similar accuracy score of 96.2% with SVM achieving a recall score of 1. Gamage and Asanka (2022) concluded that the Catboost model was highly accurate in predicting the stress level of employees and other model advanced algorithms such as deep learning and image processing can be implemented further to detect the exact mental condition from facial expressions. The gaps identified in the research were that the research was only performed on a small sample size of data and was restricted to only the IT job title.

Patil et al. (2024) discussed that use of artificial intelligence techniques in remote and hybrid work environment and how predictive analytics can help in workforce planning to create an effective remote workforce. Patil et al. (2024) discussed using quantitative data (surveys) to evaluate the productivity and satisfaction of employees and using qualitative data such as the turnover rates of employees and their attendance rates to develop a complex dataset. To identify patterns in productivity and satisfaction predictive analytics can be implemented along with regression and clustering algorithms. Patil et al. (2024) concluded that while artificial intelligence can be implemented in human resource management but research on ethical considerations and data privacy need to be considered.

2.2 Machine learning models for evaluating the productivity and satisfaction of employees.

Madan et al (2024) used 3 machine learning algorithms to measure the performance of employees. Random forest, Support vector classifier and multi layered pereptron was used to measure the performance score of the employees. The multi layered perceptron model achieved an accuracy score of 99.63% followed by Support vector classifier and Random Forest classifier. Madan et al. (2024) concluded that different parameters must be tested out for the models and the performance of the models need to be tested on different datasets. The main challenge we have identified in this study is that the accuracy score of 99.63% achieved by the multi layered perceptron model raises the concern of overfitting and the selected MLP model is a Blackbox model which doesn't provide explanations for its predictions and lacks explainability. On the contrary, Kadam et al. (2024) used 3 machine learning models to predict the future performance of employees based on their previous works. The machine learning models made use of employee performance metrics and level of education. Random forest model outperformed the other models and achieved an accuracy score of 85% followed by SVM and Decision trees. Further a feature importance analysis was conducted followed by using 10-k cross validation technique to validate the performance of the model. The study conducted by Kadam et al. (2024) demonstrated that feature importance analysis is important than high accuracies.

Mourad et al. (2022) made use of Principal component analysis for dimensionality and used multinominal logistic regression for classifying employees based on their performance. PCA was used for dimension reduction and simplifying the dataset The multinominal logistic regression model was trained to identify the performance of underperforming employees and employees with a slightly high level of performance. The multinominal achieved an accuracy score of 87.3% where the model was successful in identifying the performance of underperforming employees. However, the MNL model was unable to predict the high performing employees. Mourad et al (2022) concluded that although the model was able to predict the performance of underperforming employees it could not predict the high performing employees suggesting the need for including additional features.

Shah et al. (2024) used 6 machine learning models to investigate the impact of machine learning models on employee performance and satisfaction and identify the factors that influence performance and satisfaction. The study used machine learning models such as Random Forest, XGBoost, ADABoost, SVR and decision trees. SVR had the highest R2 score of 0.92%. In the study the model's predictions were evaluated using SHAP where the number of projects completed and salary were predicted as the top factors influencing employee satisfaction and projects completed, gender and feedback score were considered as top factors influencing employee performance.

2.3) The importance Feature Engineering Approaches in HR analytics.

Naveen (2023) presented a framework for automating feature engineering process in human resources analytics by using cloud-based ETL pipelines and advanced ml techniques. In this study advanced features were created with included the tenure of employees and performance trends. An HR based model pipeline was created and data was collected from ATS systems, LMS systems, performance management systems and engagement surveys were created for the employees. Deep feature synthesis was used to create features from Human resources data and features such as career progression of employees, employee's skill gaps were created. Techniques such as RFE (Recursive Feature Elimination) was used to select and identify best features. The importance of features was identified by using SHAP values and permutation importance. The study concluded that by using deep feature synthesis and time series feature extraction can generate actionable insights.
In the study conducted by Therasa et al (2024) used 4 machine learning classifiers to predict the overall performance of employees. A dataset of 1000 samples which consisted of employee satisfaction and productivity factors such as performance level and experience was

used. Random forest classifier, XGboost Classifier, Gaussian NB classifier and decision trees classifer were used.Random forest classifier achieved the highest performance of the 0.74 followed by XGBoost classifier, GaussianNM classifier and decision tree classifier. Therasa et al. (2024) discussed that advanced feature engineering techniques are required to improve the performance of the model, and the implementation of feature importance analysis techniques were discussed to determine which features drives the performance of the model. Therasa et al (2024) concluded that by implementing advanced feature engineering techniques the performance of the models can be increased significantly.

2.4) Limitations of the existing research.

In the above literature review, there were certain limitations that we identified. A common issue that we identified in multiple studies is the sample size selection. In the study conducted by Rehan et al. (2023) The machine learning models achieved a high accuracy, but 174 samples were collected from the employees. The high accuracy of the models over 174 samples can be misleading and the model's performance needs to be tested on new data. Okushima (2022) also collected over 400 surveys responses from various regions of Japan. Gamage and Asanka (2022) accepted that their research was performed on a small sample (481) employees and was restricted to only IT employees. Similarly, Kumara et al (2022) conducted research and gathered responses only from 250 employees in different organizations. Furthermore, in the studies conducted by Madan et al (2024) the model achieved an accuracy score of 99.63% which indicates a concern of overfitting where the model memorized data during its training instead of learning different patterns. In the study conducted by Mourad et al. (2024) the machine learning model successfully identified the performance of the underperforming employees but could not identify the high performing employees thus necessitating the need for including complex feature set. Also, most studies Also Rahman et al. (2025) in his study suggested that more work-related features need to be investigated and implemented that contribute to employee's wellbeing and productivity.

To address the above limitations in our study we selected a diverse dataset containing around 10,00 employee records. As highlighted by Therasa et al. (2024) and Rahman et al. (2025) In our study will use create advanced features that can identify the complex patterns affecting employee productivity and satisfaction metrics and try to find out to what extent does remote work have an effect on employee productivity and satisfaction metrics and what are the factors that affect employee productivity and satisfaction apart from remote work.

3 Research Methodology

This study makes use of machine learning workflow which includes steps from Data Collection, Data preprocessing, Feature engineering, Machine learning model development and Machine learning model evaluation. The dataset titled Employee Performance and Productivity data was selected from Kaggle. The dataset contains information on performance, satisfaction of employees, the frequency of remote work by employees and the number of projects handled by employees. The selected employee performance and productivity data will help us in determining whether if remote work has an impact on the efficiency and productivity of employees.

3.1 Dataset Description

The Employee Performance and Productivity Data was selected from Kaggle an open source an open-source platform. The dataset contains 10,000 unique employee records and employee details as Gender, Age, Job Title, Monthly salary, performance score, satisfaction score, hours worked per week, sick days, remote work frequency, promotions and project handled by employees.

```
Employee_ID
Department
Gender
Age
Job_Title
Hire_Date
Years_At_Company
Education_Level
Performance_Score
Monthly_Salary
Work_Hours_Per_Week
Projects_Handled
Overtime_Hours
Sick_Days
Remote_Work_Frequency
Team_Size
Training_Hours
Promotions
Employee_Satisfaction_Score
Resigned
```

Fig 1: -Attributes in Employee Productivity and performance dataset

3.2 Data preprocessing

Data preprocessing an important step before analysing the dataset the selected Employee performance and productivity dataset went through several data preprocessing steps to ensure data quality and feasibility. Data preprocessing is very important for making sure that the data we are going to use for modelling and analysis is clean and has no noise. The first step conducted in data preprocessing was to ensure that no missing values were already present in our selected dataset. Upon inspecting the dataset there were no missing values in the dataset and the dataset contained 10000 unique rows and 20 column features. The Dataset consisted of both numerical and categorical features. Numerical features included performance score, satisfaction score, remote work frequency, projects handled and sick days. The categorical features in the dataset included education, job title, department and gender. The key variable Remote work frequency consisted of values ranging from 0,25,50,75,100 representing the employee's level of remote work where 0 implies that the employees work fully from office and 100 implies fully remote work employees.

<u>Outlier Detection and handling</u>

After inspecting the dataset and ensuring that the data is complete, the next step was to find out if there were any potential outliers present in the data. Numerical columns which included age, Years_at_company, Performance_Score, Work_Hours_Per_Week, Monthly_Salary and Projects_Handled were checked for potential outliers. Outliers were detected in work_hours_per_week column The work_hours_per_week column consisted of potential outliers as we found that many entries reflected unusually high working hours. Further investigation revealed that the work_hours_per_week column had extreme values exceedingly above 40 hours, a working week schedule. The work_hours_per_week column was capped around 40 hours a week to reduce the impact of extreme values on the performance of machine learning models and to preserve the integrity of the data and to prevent skewness. Value clipping technique was used to cap the upper bound values of the work_hours_per_week column to 40 hours a week to ensure a standard working schedule of full-time employees. After handling the outliers in the hours_per_week column outlier detection was carried out on key numerical columns which included Age, Years_at_company, Perfomance_score, Remote_work_frequency and satisfaction_score. For detecting the outliers in numerical columns Interquartile range method (IQR) was used. IQR is calculated as [IQR] = Q3-Q1 The Interquartile Range method (IQR) revealed that no outliers were present in Age and had an actual range from 22 to 60 and a valid range of 1 to 81, Remote_work_frequency had no outliers and an actual range from 0 to 100 and a valid range of -50.00 to 150.00, Years_at_company had no outliers and an actual range from 0 to 10 years and valid range from -5.50 to 14.50 years. Similarly, no outliers were present in the performance score and Satisfaction score, and they had an actual range of 1 to 5 scale. In the outlier detection analysis outliers present in hours_per_week column were handled using value clipping method and were capped to 40 hours a week to represent a typical working week schedule of employees and Interquartile Range method IQR was further used to detect outliers in key columns Age, Years_at_company, Remote_work_frequency, performance_score, satisfaction score. The Interquartile range method revealed that no outliers were present in the key columns, and the dataset was reliable and suitable for training and developing machine learning models.

Exploratory Data Analysis.

After detecting and handling the outliers a detailed exploratory data analysis was carried out to identify complex patterns and different relationships between variables. The exploratory data analysis phase provided useful insights about the distribution of key variables such as remote work frequency and employee metrics such as hours_worked_per_week, monthly_salary, Performance score, satisfaction score. First a data validation technique was used for age and years of experience columns to check if there were impossible age and years at company combinations. The data-validation technique revealed 3784 invalid records of age and years at company combinations for e.g. employees aged 23 having 9 years of experience is impossible assuming the legal age of working fulltime in most of the organizations is 18 years.

```
Invalid records: 3784
Invalid age and years of experience combinations exist:
     Age  Years_At_Company
19    23                 9
81    23                 7
171   22                 9
217   22                 6
226   25                 9
```

Fig 2: -3784 invalid records of age and years at company

To solve the issue of impossible age combinations the maximum possible experience for employees was calculated where the minimum age to start working was set to 18 years. Then the years at company were capped to feasible values using np. minimum function and aligned with the set minimum working age i.e. 18 years to ensure that no impossible age and years at company combinations exist. After ensuring that no impossible age combinations exist in the data, the distribution of remote_work_frequency was analyzed. The remote_work_frequency distribution was examined to find out different working patterns of employees. The range of remote_work_frequency distribution was from 0,25,50,75,100. For further analysis the remote work frequency was distributed into 5 different working models where 0% remote represented office-based employees, 25% represented minimum remote working employees, 50% represented hybrid working employees, 75% represented mostly remote working employees and 100% represented fully remote working employees. The distribution of different working models was uniform where there were 19837 employees working from office, 20235 employees doing minimum remote work, 19706 hybrid employees. The uniform distribution of different work models made sure that there was no bias towards any work model. After distribution of remote_work_frequency into 5 different work model categories key features as performance score, satisfaction score, hours_worked_per_week, projects_worked were compared across different working models.

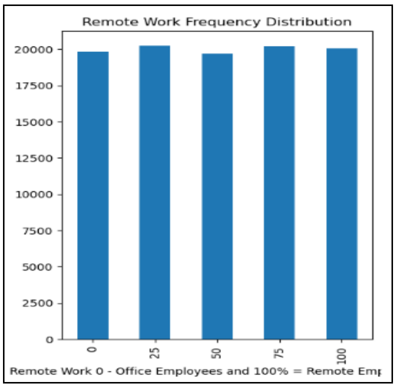

Fig 3: - Distribution of remote_work_frequency

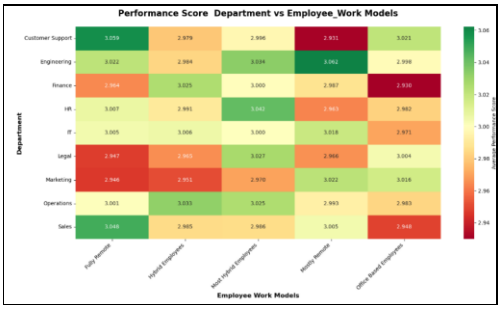

Fig 4: - Heatmap of performance score department vs employee work models.

After analysing the distribution of remote_work_frequency a heatmap was plotted to analyse performance score of different departments across different employee working models. The heatmap visualization. The performance score scale has a rating scale from 1 to 5. The heatmap visualization indicated that performance score was consistent across various departments and different employee working models with values ranging from 2.946 to 3.062. Employees in the Engineering department had the highest performance score of 3.062 and employees from IT department had a consistent performance score over all working models. Employees from the customer support department had the highest performance score through different working models where employees working fully remote had a performance score of 3.059 and employees working from the office had a performance score of 3.021. The analysis indicated that employees working in customer department roles were different in terms of productivity and efficiency across different working environments. On the other hand, employees working in finance department preferred hybrid working arrangements and had high performance scores of (3.025) in hybrid working models over remote work model and office-based work model. The heatmap visualization provided insights that the performance score of all the departments across different working models remains consistent, however different work preference patterns exist for some departments for instance employees working in customer support department benefited in office based working model possibly due to collaboration and few distractions. Meanwhile, employees working in the engineering department benefited in mostly remote working environments.

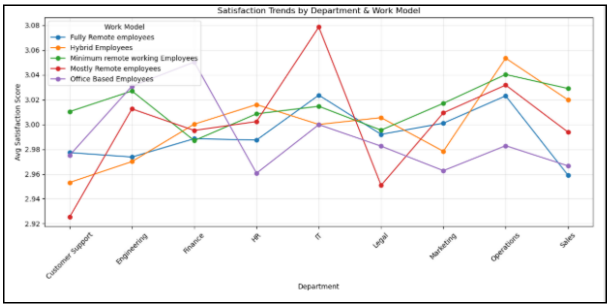

Fig 5: -line chart of satisfaction score of departments across different work models.

The Line Chart visualization provided insights on employee's satisfaction score across different working models. Employee satisfaction score had a range from 2.92 lowest to 3.08 highest. Mostly remote employees working in IT department had the highest satisfaction score of 3.08, signifying their adaptability to remote work culture. On the contrary employees working mostly remote and in hybrid working environments had a low satisfaction score of 2.92 and 2.94 respectively while office-based employees reported higher satisfaction scores. Employees in Engineering department had consistent satisfaction scores across different work models (2.98 to 3.04) signifying that the department was adaptable to different work models. Office based employees in the finance department reported the highest satisfaction score of 3.04 and the employees working in hybrid work model in the operations department reported high satisfaction scores. Marketing departments displayed interesting insights that employees working fully remote reported low satisfaction score (3.00) than the employees mostly working remote (3.02) suggesting that employees in marketing department don't benefit from fully remote work model but benefit from minimum remote work model suggesting the need of hybrid working model while ensuring flexibility in remote work model, The Line Chart visualization of Employees satisfaction score provides insights that employee satisfaction is highly dependent on different departments and job functions. While from the heatmap visualization performance score was consistent across the different working models but the employee's satisfaction line chart provided insights that satisfaction of employees in different departments is more dependent on their working arrangements suggesting that organizations need to implement flexible working arrangements policies for employees in different departments based on their satisfaction level to ensure maximum efficiency and productictivity.

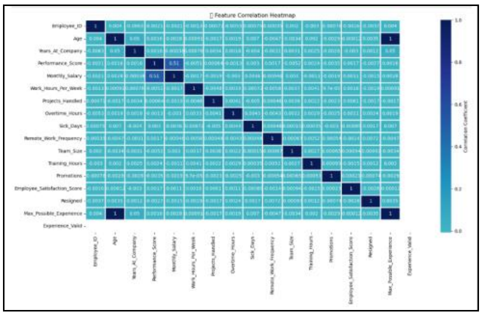

Fig 6: - Correlation Matrix Visualization displaying weak correlations

A correlation matrix was plotted to find meaningful relationships between different variables. A strong relationship was identified between monthly salary and performance score as higher employee performance results in higher the salary Age and years at company have a moderate correlation indicating that age of employees increases with years spent at company. Remote_work_frequency didn't have strong insights with other key variables and with variables like performance_score and Employee satisfaction score. Also, Employee performance score and Employee satisfaction score had a weak relation The weak correlation of remote_work_frequency with key variables like performance_score and Employee satisfaction score indicates that it is affected by different factors thus creating a need for feature engineering. The correlation matrix indicated that no strong relationships existed between the key variables and cited the need for creating complex and advanced features in feature engineering phase.

3.3 Feature Engineering

Feature Engineering was used to create new features from our existing variables in the dataset to capture the complex relationship between remote_work_frequency and employee productivity and efficiency.

Employee Efficiency Score was created from employee performance score and employee satisfaction score. The Employee Efficiency Score created from employee performance score, and employee satisfaction score is a complex feature and a target variable which will be used for prediction. The composite Employee efficiency score makes use of both employee performance and satisfaction over traditional productivity metrics which only considers the performance of employees. While creating the employee efficiency score employee performance score was assigned a weightage of 70 and employee satisfaction score was assigned a weightage of 30 as the productivity of employees require a balance between their performance and satisfaction for better results. The approach used for creating Employee Efficiency score ensures that employee satisfaction and wellbeing are also considered with their performance for better results. The performance score and employee satisfaction were normalized and converted to a rating scale of 1 to 5 with distribution indicating multiple peaks at 2.0, 2.5, 3.5, 4.0. The mean efficiency score of 3.0 indicates a more efficient, productive and satisfied workforce.

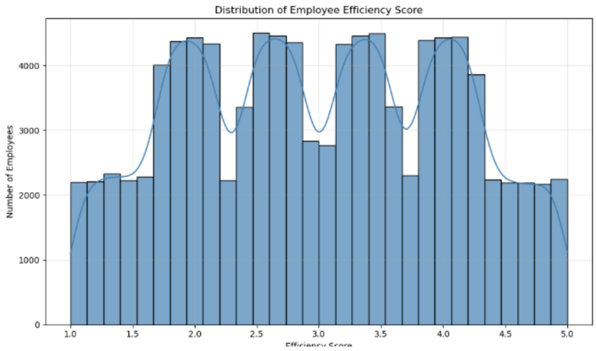

Fig 7: - Distribution of Employee efficiency score

Employee Workload level was created by combining the regular hours employees work in a week and the overtime hours that employees work. The Employee workload level was created to find out the effect of hours that employees work for have on productivity and efficiency of employees. The total and overtime hours that employees work for were grouped into 3 categories where standard hours included the typical hours that the employees worked in an organization (40) hours a week, High hours if the employees worked for 50 50 hours and extreme hours if the employees worked more than 50 hours. The Employee workload level was created to find out the effect of hours worked by employees on productivity and efficiency of employees.

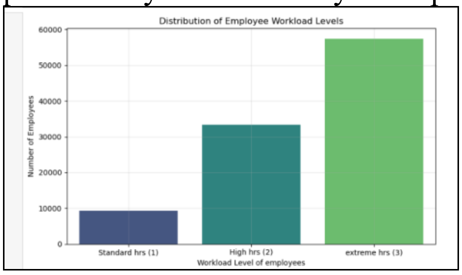

Fig 7: - Distribution of Employee Workload level

Employee Career Progression was created by combining years at company and number of promotions to measure the progression of employee careers. The composite Employee Career progression feature captured the growth pattern of employees in different departments in the organization. The distribution of employee career progression was balanced indicating balanced career growth of employees in the organization.

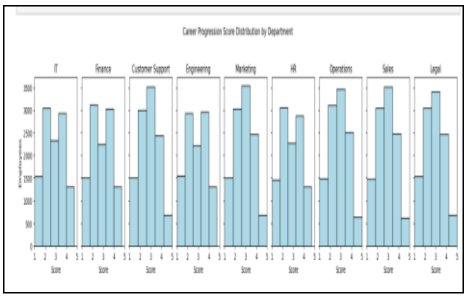

Fig 8: - Distribution of Employee Career Progression

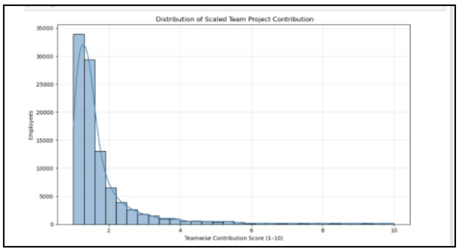

Fig 9: - Distribution of projects handled teamwise score skewed

Projects handled teamwise score feature explained the contributions of employees in different departments to projects on a scale of 1-10. Projects handled teamwise score was created by dividing the number of projects handled by employees with the size of their teams. The original distribution of Projects handled teamwise score was highly skewed on the right side where the distribution had a long tail towards right side with high scores. To address the skewness issue Yeo-Johnson power transformation was used. Yeo-Johnson transformation a data transformation technique used to handle skewness and transform data into a normal distribution. Applying Yeo-Johnson power transformation resulted in a normalized distribution of data. As per Fig 9, the Yeo-Johnson transformed data represented a bell-shaped distribution.

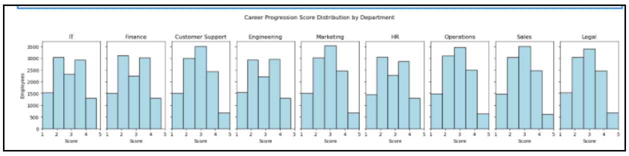

Fig 9: - Distribution of projects handled teamwise score normalized

As the correlation analysis in Figure 6 displayed that remote_work_frequency had no strong relationships with the key variables and suggested that remote work doesn't affect every employee in a same way so 2 remote work interaction features were feature engineered1) RemoteEmployees_TeamContribution_Interaction
2) RemoteEmployees_Careerprogression_Interaction

RemoteEmployees_TeamContribution_Interaction feature was created by combining remote work frequency and Teamwise_Contribution_Score The main aim of the RemoteEmployees_TeamContribution feature was to find if employees working in high remote frequencies productive and efficient and contribute to their teams irrespective of their worklocation.

RemoteEmployees_Careerprogression_Interaction was created by combining Remote work frequency and Employee_Career_Progression The main purpose of creating RemoteEmployees_Careerprogression_Interaction was to find out if working remote has an effect on the career progression of employees and to find how well do employees thrive in remote working environments.

Employee_Salary_ToYears ratio was engineered by dividing monthly salary of employees by the years they worked at the company (+1 for new employees) The Employee_Salary_Toyears ratio was engineered to identify high performance employees who received a quick growth in salary progression.

7 new features including the target variable were created from existing feature set which provided insights and would be used for training machine learning models to find out if remote work has impact on productivity and efficiency.

3.4 Machine Learning Model Selection

After Feature Engineering phase, 2 machine learning models were selected and developed to first the employee performance and productivity was divided into training and testing dataset 80% data was allocated for training the models and 20% data was used for testing the models.2 machine learning models were selected for analysing the impact of remote work on our newly feature engineered target variable Employee Efficiency Score.

3.5 Machine Learning Model Evaluation

For evaluating the performance of models, evaluation metrics such as R2, mean absolute error (MAE) and mean squared error were used. R2 is called coefficient of determination was selected as primary performance evaluation metric as it displays the level of variance explained by models where values that are closer to 1 indicate that the implemented models have a good prediction power. Mean square error MSE was selected as it measures how far are the predictions of the model from the true values and mean absolute error was selected because as it explains the difference between actual values and the predicted values. The 3-evaluation metrics were selected and applied to both Random Forest Regressor and Light gradient Boosting machine models to determine how accurate are the predictions made by both the models.

4 Design Specification

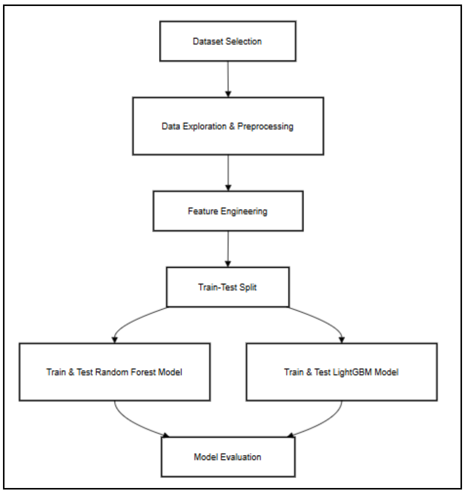

Fig 10: - System Framework of our proposed system

Fig 10 displays the framework of the designed system to find out if remote work has effect on employee's efficiency and productivity. The framework designed has components starting from selection of dataset, Data exploration and preprocessing, Feature Engineering, Dividing the data into training and testing splits, Training and testing the regression models and concluding with evaluating the performance of the models.

4.1 Machine Learning Model selection.

Random forest and Light Gradient Boosting machine are regression models that were selected. Random forest model was selected as a baseline model and Light Gradient Boosting machine LGBM was selected as an advanced model. These both models were selected because of their ability to handle predictions with non-linear data.

Random Forest

Random forest model is a bagging model and a group of decision trees where each tree is trained separately on a subset different from its data and a new subset is obtained with the help of bootstrapping.

Random forest is a machine learning model that uses multiple decision trees to create predictions. The reason behind selecting Random Forest model was that the model is prone to outliers and overfitting and can handle nonlinear relationships naturally and provides feature importance analysis where the features that drive the behaviour of the model are identified.

Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine is a boosting framework and is recognized for its performance. LGBM is a tree-based model that builds multiple decision trees sequentially where each new trees learns from the mistakes and errors from the previous decision trees and makes use of a leaf wise strategy instead of a level wise strategy which makes increases its accuracy and predictions making it superior to other Gradient Boosting models.

5 Implementation

The implementation chapter discusses the tools that were used to develop the machine learning system framework which consisted of different phases right from Dataset selection phase to Model development and evaluation phase for analysing the impact of remote work on employee efficiency and productivity. Python 3.12.4 is used as the main programming language and Jupyter notebook as the main development environment. Different libraries imported for data manipulation, analysis and visualization were pandas, NumPy, seaborn and matplotlib. Pandas was used for finding out and handling missing values, duplicate records and for feature engineering. NumPy was used for numerical operations especially while creating our target variable Employee Efficiency Score. Matplotlib and seaborn were used for visualizations and plotting distributions of the key variables in our dataset and exploring different types of correlations present in our data with the help of correlation matrix and heatmaps.

Feature selection and Data splitting for model training

The Employee productivity and performance was divided split into training and testing sets where 8000 samples were allocated for training the models and 2000 samples were allocated for testing. The splitting was done with the help of Sckit-learn's train_test_split function, and the random state was set to 42 to ensure reproducibility. To prevent overfitting and multicollinearity 7 key features were selected. The selected 7 key features were

1) Employee_Workload_Level
2) Employees_Salary_to_Years_worked_ratio
3) Remote_Employees_CareerProgression
4) Remote_Employees__TeamContribution
5) Remote_Work_Frequency
6) Projects handled teamwise score
7) Employee career progression

The 7 key features were selected based to prevent data leakage and overfitting The other features were not selected to prevent redundancy and multi collinearity while including features that are critical to understand the impact of remote work.

Machine Learning Model Selection.

2 Machine learning models were selected for our study. Random forest Model was selected as a Baseline model because it can handle complex non-linear relationships and its ability to provide insights of the key features that drive its predictions. Random forest model also makes use of multiple decision trees and combines the predictions of multiple decision trees to create a final prediction. The parameters of random forest model are n estimators which means the number of decision trees, max depth which means the maximum level of growth of the decision trees. The Random Forest Model was selected for our study because our dataset contained nonlinear relationships between remote work frequency and employee productivity and satisfaction metrics.

Light Gradient Boosting Model a high-performance gradient boosting framework was selected as the advanced model for our study. LGBM is a tree-based model that builds multiple decision trees sequentially where each new trees learns from the mistakes and errors from the previous decision trees. In contrast to Random Forest model in which decision trees are build parallel, LGBM builds decision trees one after another in which the new tree learns from the mistakes of the previous tree. LGBM uses a leaf wise strategy instead of a level wise strategy which makes increases its accuracy and predictions making it superior to other Gradient Boosting models. The parameters of LGBM include n-estimators which means the number of decision trees to build for our study we set the number of estimators to 250 to risk overfitting, learning-rate which means the speed at which the new decision tree learn from the errors of the previous decision trees. The learning-rate set was set to 0.6, max-depth which means the level of growth of single decision tree which means how far a single decision tree can grow the max-depth was set to 7 because the deeper the decision trees grow more patterns are identified and captured. LGBM was selected for our study because of its ability to handle tabular data and superiority in identifying complex patterns between remote work and productivity and satisfaction metrics.

Results

Both the models, Random Forest and Light Gradient Boosting Machine model were trained on 8000 samples of data, and their performance was evaluated on 2000 samples of data. First random forest model was implemented with default parameters where the parameters were not tuned on the test data. The model achieved an R2 score of 0.3103, where the model explained 31.03 of its variances, mean absolute error of 0.71 where the predicted values differed by 0.71 units from the actual values and Mean squared Error of 0.76 a square difference between the actual values and the predicted values.

To Further improve the prediction power of Random Forest model, we conducted hyperparameter tuning using GridSearchCV with Cross-validation. During hyperparameter tuning new parameters along with the default parameters were added to the parameter grid which included min_sample_split, min_leaf_split and max_features. By testing and trying different parameters combinations in hyper parameter tuning the random forest model was optimized and achieved an R2 score of 0.5175 where the model was able to explain 51.75 % of its variance, MAE of 0.58 and MSE of 0.53. By using hyperparameter tuning along with Grid search and Cross validation the R2 score of the model increased from R^2: 0.3103 to 0.5175.

After implementing and hyperparameter tunning of the Random Forest model LightGradient Boosting model was implemented on the test data. New parameters were added along with the default parameters were added to the parameter grid. Regularization parameters L1 and L2 were applied to the model. L1 regularization parameter was added which promotes sparsity and prevents overfitting and L2 Regularization parameter was added which helps in shrinking weights to improve generalization Further performance specific parameter called force-row-wise was added which forces the LGBM model to build histograms row-wise instead of column wise. After setting the parameters, the LGBM model was implemented. The Light Gradient Boosting Model achieved an R2 score of 0.8099, MAE score of 0,36 and MSE Score of 0.21. The performance of LGBM model outperformed the tuned Random Forest Model over all the metrics. The exceptional performance of LGBM model over Random Forest model indicated that GBM approach is very effective in identifying complex relationships between remote work and employee efficiency.

6 Evaluation

In the Evaluation chapter, the model's performance is discussed also feature importance analysis in conducted in this section to find the prominent features driving the decisions of the models.

Evaluation metrics such as R2, Mean absolute error and mean square error are used here to evaluate the performance of the Models.

6.1 Experiment 1 Random Forest Model Evaluation.

A Random Forest model selected as the baseline model and was implemented with default parameters to determine the impact of remote work on employee's efficiency. Root square (R2), Mean absolute error (MAE) and mean square error (MSE) were the performance metrics used to evaluate the performance of the Random Forest Model. The model achieved an R2 score of 0.3103 where the model explained 31.03% of variance in employee efficiency scores. MAE score of 0.71 and MSE score of 0.76 suggesting that the prediction power of the model was moderate and there was room for improving the performance of the model.

```
 Baseline Random Forest:
R²: 0.3103
MAE: 0.71
MSE: 0.76
```

Fig 11: - Performance of Baseline Random Forest Model.

. . .

6.2 Experiment 2 Hyper Tuned Random Forest Model Evaluation.

After the initial implementation of Random Forest Model where the model achieved an R2 score of 0.3103% The parameters of the Random Forest model were hyper tuned using Grid search CV and 3k cross validation. After tuning the parameters of the Random Forest model, the performance of the model increased significantly resulting in an R2 score of 0.5175, MAE 0.58 and MSE 0.53 which reflects that the prediction power of the model increased significantly.

```
 Hyper Tuned Random Forest:
R²: 0.5175
MAE: 0.58
MSE: 0.53
```

Fig 12: - Performance of Hyperparameter tuned Random Forest Model.

After evaluating the performance of the hyper tuned random forest model, feature importance analysis was conducted to determine which key variables drive model's decisions. The feature analysis revealed the features the ranking of important features.

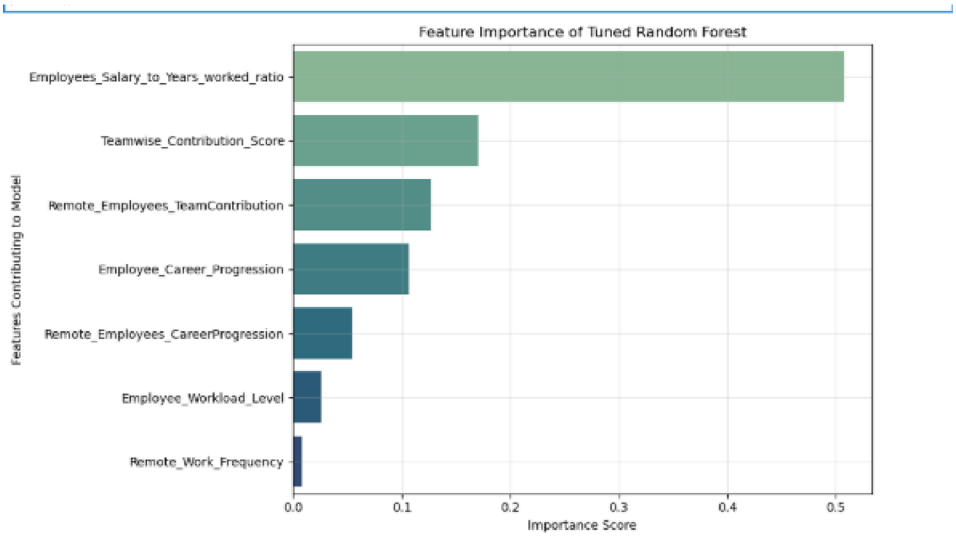

Fig 13: - Feature importance of Hyperparameter tuned Random Forest Model.

The feature importance analysis that the most important feature for employee efficiency was Employee salary to years ratio, Employees Projects handled score followed by Remote_Employees_Team_Contribution (Remote work interaction feature), Employee Career_progression, Remote_Employees_Career_progression,(Remote work interaction feature), Employee_Workload_level and Remote work frequency was considered lowest important feature Remote_Employee_Career_Progression and Remote_Employees_Team_Contribution contributed more to the model's prediction power rather than remote work frequency. After finding out the important features that drive model predictions A Predicted vs actual values scatter plot was plotted which displayed a strong linear relationship between the predicted employee efficiency score vs the actual employee efficiency score. The Tuned Random Forest model displayed consistent performance over the entire efficiency score range (1.0 to 5.0) with many predictions falling closely towards actual values. The tight distribution of points towards the diagonal line confirms that the tuned random forest model was successful in identifying the factors affecting the employee efficiency scores. Minor scatters were identified in the mid-range of employee efficiency score (2.0 to 4.5). The strong correlation between the actual values and predicted values confirms that our feature engineering approach, specifically the remote work interactions engineered were successful in capturing complex patterns in employee's efficiency. The random forest model prioritized the remote work interactions over direct remote_work_frequency which indicated that employee's efficiency depends on how remote work is implemented and incorporated with career development of employees and team collaboration.

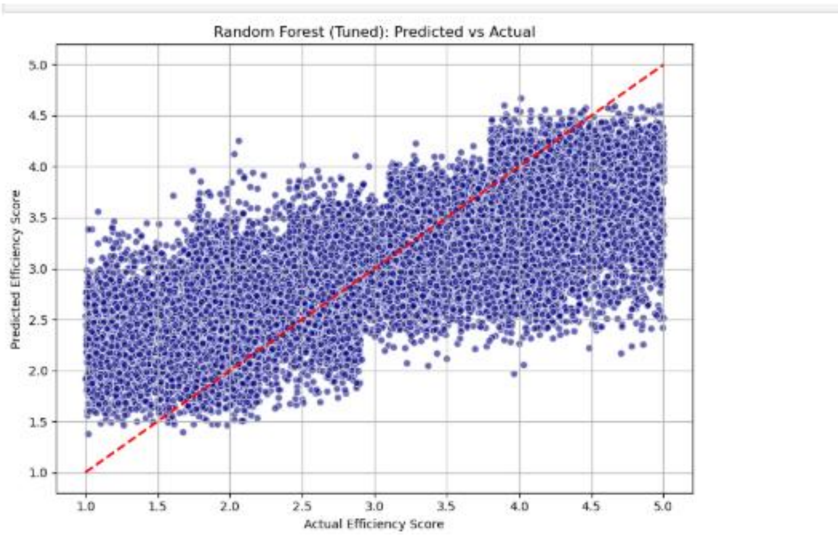

Fig 14: - predicted vs actual values tuned Random Forest Model.

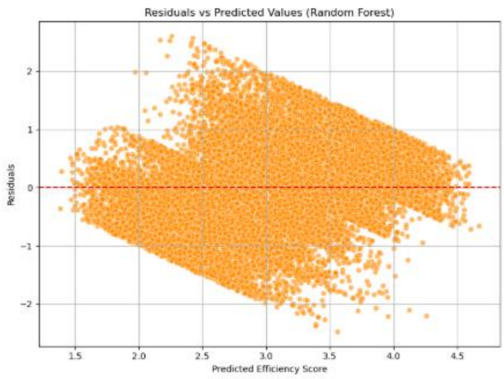

Fig 15: - predicted vs residual values of tuned Random Forest Model.

6.3 Experiment Light Gradient Boosting Model Evaluation.

After Random Forest model Light Gradient Boosting Model was selected as the advanced model and was implemented to determine the impact of remote work on employee's efficiency. Before implementing the LGBM model its parameters were carefully hyper tuned and Regularization parameters L1 and L2 were applied to the model. L1 regularization parameter was added which promotes sparsity and prevents overfitting and L2 Regularization parameter was added to improve generalization. The boosting type was selected as gbdt (Gradient Boosting decision tree method) n-estimators (no of trees) were set to 250 as more with trees training time increases and there is also a risk of overfitting. Learning rate was set to 0.6 and maximum depth of the trees was set to 7. Subsample which means a portion of training data allocated during boosting was set to 0.5. regularization parameters reg_alpha, reg_lamda were used to add stability and prevent overfitting and force_row_wise parameter was set to true so that the LBGM model created histograms row wise instead of column wise histograms and is useful for improving the performance. After carefully hyper tuning the parameters of LGBM model it was implemented.

```
lgbm_model = LGBMRegressor(
    boosting_type='gbdt',
    n_estimators= 250,
    learning_rate=0.6,
    max_depth=7,
    subsample=0.5,
    colsample_bytree=0.5,
    random_state=42,
     reg_alpha=0.8,
    reg_lambda=0.8,
    force_row_wise = True,
```
Fig 16: - Parameters of LGBM.

The LGBM model achieved an R2 score of 0.8099 where the model explained 80.99% of variances in employee efficiency score. MAE of 0.36 and MSE of 0.21. The LGBM model implemented had a high performance over Tuned random forest model in all evaluation metrics.

```
[LightGBM] [Info] Total Bins 879
[LightGBM] [Info] Number of data points in the train set: 80000, number of used features: 7
[LightGBM] [Info] Start training from score 2.999730
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
R²: 0.8099
MAE: 0.36
MSE: 0.21
```
Fig 17: - Performance of Light Gradient Boosting Model.

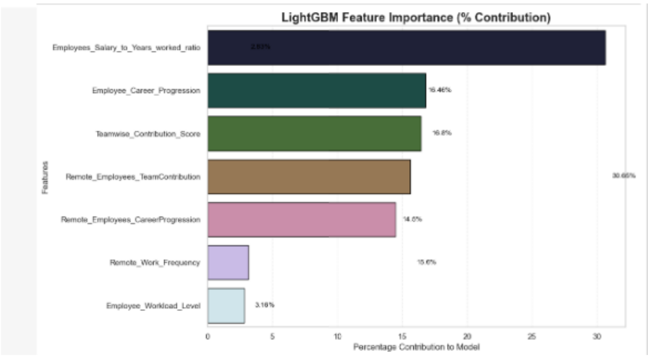

Fig 18: - Performance of Light Gradient Boosting Model.

	Feature	Importance (%)
3	Employees_Salary_to_Years_worked_ratio	30.65
2	Employee_Career_Progression	16.80
1	Teamwise_Contribution_Score	16.46
5	Remote_Employees_TeamContribution	15.60
4	Remote_Employees_CareerProgression	14.50
6	Remote_Work_Frequency	3.16
0	Employee_Workload_Level	2.83

After evaluating the performance of LGBM a feature importance analysis was performed where the model considers determining which key variables drive its decisions. The important features were Employee_Salary_to_years_ratio which has a contribution of 30.65%, followed by employee career progression 16.80% and Employees Projects handled score at 16.46% The remote work interaction features that were engineered displayed significant influence where the Remote_Employees_Teamcontribution contributed 15.60% and Remote_Employees_Careerprogression contributed 14.50% whereas remote_work_frequency contributed only 3.16% followed by employee_workload_level was the last feature which contributed 2.83%. The remote interaction terms feature engineered when combined contributed around 30.10% towards model's performance over remote_work_frequency which only contributed 3.16%. The feature analysis of LGBM had similar patterns when compared to the tuned random forest model but had a more balanced feature importance distribution.

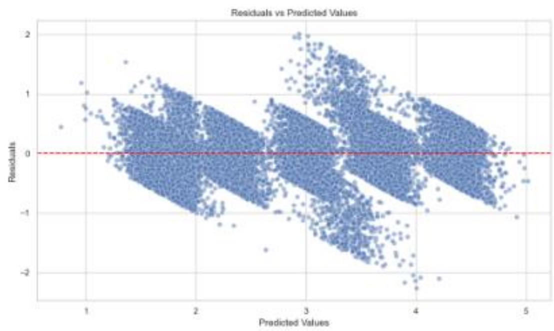

Fig 19: - Residuals vs predicted values of LGBM

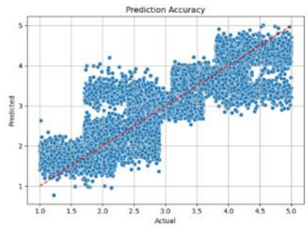

Fig 20: - Prediction accuracy of LGBM

After feature importance analysis, A scatter plot of model's prediction accuracy was plotted. The plot revealed the model's prediction accuracy where the predicted values of the model were in a near perfect relationship with the actual values of efficiency score. The data points were tightly clustered at the diagonal line which indicated that the LGBM model had accurate predictions towards the employee efficiency score (1.0 to 5.0) with minimum scatter around the red line. The random scatter pattern indicated that no bias is present in the model, and its performance is consistent across all the predicted values. Also, the constant distribution of residual value over employee efficiency score (1.0 to 5.0) shows homoscedasticity which means that the model's predictions remain consistent irrespective of the employee efficiency score that is predicted. A strong correlation is observed in the predicted vs actual scatter plot which indicated that our feature engineering approach especially the remote work interaction features were successful in capturing the complex patterns influencing employee efficiency scores.

Comparative Analysis of Models.

Models	R2 Score	MAE (Mean absolute Error)	MSE (Mean Square error)	Variance explained by model
Random Forest Default	0.3103	0.71	0.76	31.03%
Tuned Random Forest Model	0.5175%	0.58	0.53	51.75%
Light Gradient Boosting Machine (LGBM)	0.8099	0.36	0.21	80.99%

Fig 21: - Comparative analysis of all the models.

7 Discussion

The results of our study conducted provided evidence that the impact of remote work doesn't have a direct strong influence on employee efficiency but has effect through interaction features. In our study remote work frequency contributed (3.16%) towards the LGBM's predictive power but the remote work interaction features Remote employees team contribution and remote employees career progression contributed 30.10% when combined. This finding in our study indicated that the effects of remote work are complex and not only determined by its frequency but by how remote work is implemented in collaboration between teams and the support they receive in their career journey.

In our study, the most important factor having an impact on employee's efficiency was Employee_salary_to_years_ratio which solely contributed 30.65% towards employee efficiency which means that salary progression acts as an important factor of employee performance. Employees with high salary growth display high productivity and satisfaction and are motivated, possessing strong professional values which make them highly efficient irrespective of their work location. This finding indicates that organizations must focus more on the career development and progression of employees rather than focusing more on the working environments. The LBGM model we used in our study outperformed the tuned random forest model which indicated that Gradient boosting models are highly effective in uncovering the complex patterns between remote work and employee efficiency.

Limitations of our Study

There are several limitations in our study that we need to be acknowledge. Despite the LGBM model having a high accuracy score of 0.8099 we only selected 7 complex features which were we feature engineered. The reason for selecting 7 feature engineered features was to prevent model overfitting and avoiding multi collinearity. We didn't select features such as age, education level and department during model training and testing they would have provided further insights if they would have been encoded properly. Also our feature engineering approach in creating interaction features was limited to only 2 features Remote employees career progression and remote employees team contribution we could have feature engineered more interaction variables such as remote work department wise and remote work with years of experience, remote work sick days Also we could have included original features in our feature set such as projects_handled, work_hours and sick_days to capture the raw effects on employee efficiency these features were not selected due the risk of overfitting and in order to ensure model simplicity. We only used random forest and LGBM other boosting models such as XGBoost, ADABoost could have been used also deep learning approaches which could have helped in achieving better results. Also, our feature engineering was highly dependent on knowledge of HR domain, and we could have used more advanced feature selection approaches such as lasso and RFE to identify the ideal feature sets. Also, the target variable we created employee efficiency score which was assigned weights of 70-30 (70 for performance and 30 for satisfaction) the weighting logic was based on the practices of organizations where the performance is given a high preference and satisfaction is given a second preference. However, that changes from organization to organization where satisfaction is also given an equal preference with performance. Lastly, we also needed to test the performance of the LGBM model on multiple datasets to ensure consistent performance. These were the limitations present in our study that we acknowledge.

8 Conclusion and Future Work

Our study's aim was to answer the question does remote work has a direct impact on the performance and satisfaction of employees. After extensive analysis, feature engineering and development of ml models we found that the impact of remote work on employee performance and satisfaction metrics is indirect, and remote work affects employee performance and satisfaction through complex patterns and remote work is highly dependent on individual characteristics of employees such as their career progression and their collaboration between team. The important factor that affects employee productivity and satisfaction is Employees_Salary_to_years_ratio. For future work on finding the impact of remote work on employee productivity and satisfaction metrics time series analysis can be conducted to capture the effects of remote work on performance and satisfaction metrics over time which can capture and identify seasonal patterns of performance and satisfaction of employees. Also, the features related to communication, task completion rates and projects timelines could be included to provide a more accurate understanding of remote work's impact on productivity and satisfaction metrics. As most of the organizations are implementing remote work the above approaches will be useful for maximizing productivity and satisfaction.

References

Rahman, Md. W., Anjum, I., Hamim Mahi, S. A., Faisal Karim, M., Ahad, A., & Imam Bijoy, Md. H. (2025). Machine Learning-based Analyzing Impact of Remote Work Arrangements on Employee Well-Being and Productivity. *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6. https://doi.org/10.1109/ECCE64574.2025.11013806

Jacob, J., & Jayalakshmi, G. (2025). Analyzing the Effects of Remote Work on Productivity and Work-Life Harmony using Federated Learning Models. *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 1–7. https://doi.org/10.1109/ICDCECE65353.2025.11034817

Okushima, M. (2022). Application of machine learning models for partial remote work intension in provincial areas of Japan. *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, 1–6. https://doi.org/10.1109/SCISISIS55246.2022.10002154

Elshaiekh, N. E. M., Hassan, Y. A. A., & Abdallah, A. A. A. (2018). The Impacts of Remote Working on Workers Performance. *2018 International Arab Conference on Information Technology (ACIT)*, 1–5. https://doi.org/10.1109/ACIT.2018.8672704

Kumara, B. T. G. S., Herath, G. A. C. A., Wijeratne, P. M. A. K., & Banujan, K. (2022). Work From Home After Covid-19: Machine Learning-Based Approach to Predict Employee's Choice. *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 147–150. https://doi.org/10.1109/DASA54658.2022.9765249

Jaja, J., Priatna, N., & Ardan, T. S. (2021). Implementation of Data Mining Technique for Performance of WFH and WFO Agents Using the K-Means Method Case Study Study of PT. Infomedia Telkom Consumer Profiling Services. *Budapest International Research in Exact Sciences (BirEx) Journal*, *3*(2), 117–125. https://doi.org/10.33258/birex.v3i2.1810

Sahyaja, C., Shankar, C., Zeeshan, K., & Nagaraj, P. (2024). Artificial Intelligence and Remote Work: Transforming Human Resource Management in a Post-Pandemic World. *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 1555–1560. https://doi.org/10.1109/ICACRS62842.2024.10841784

Hammoumi, A. el, Seghyar, N., Hammoumi, A. M. el, & Ouali, A. el. (2024). The Impact of Artificial Intelligence on Employee Adaptability to Remote Work. *2024 3rd International Conference on Embedded Systems and Artificial Intelligence (ESAI)*, 1–5. https://doi.org/10.1109/ESAI62891.2024.10913897

Patil, S., Upadhye, P., Khandekar, S. R., Darekar, P., Ghewari, A., & Bhattacharya, U. (2024). Artificial Intelligence and Data Science Approaches to Analyzing Human Resource Management in Hybrid and Remote Work Models. *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 1–6. https://doi.org/10.1109/ICSES63760.2024.10910810

Rehan, F. A., Bukhari, F., & Iqbal, W. (2023). Impact of COVID-19 on Productivity of Software Engineers: A Comparative Analysis of work from home (WFH) and work from office (WFO) environment using Machine Learning. *2023 2nd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, 1–8. https://doi.org/10.1109/ETECTE59617.2023.10396746

Elshaiekh, N. E. M., Hassan, Y. A. A., & Abdallah, A. A. A. (2018). The Impacts of Remote Working on Workers Performance. *2018 International Arab Conference on Information Technology (ACIT)*, 1–5. https://doi.org/10.1109/ACIT.2018.8672704

Nilushika Gamage, S., & Dinesh Asanka, P. P. G. (2022). Machine Learning Approach to Predict Mental Distress of IT Workforce in Remote Working Environments. *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 211–216. https://doi.org/10.1109/SCSE56529.2022.9905229

Madaan, V., Singh, C., Jain, I. B., Wadhwa, S., Sarangi, P. K., & Sahoo, A. K. (2024). Employee Performance Measurement Using Machine Learning Algorithms. *2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG)*, 1–5. https://doi.org/10.1109/ICTBIG64922.2024.10911748

Manoharan, G., Perla, S., Kadam, P. S., Sathe, Dr. M., Ranichandra, S., & Lourens, M. (2024). Supervised Learning in Employee Performance Prediction. *2024 Asian Conference on Intelligent Technologies (ACOIT)*, 1–5. https://doi.org/10.1109/ACOIT62457.2024.10940050

Mourad, Z., Noura, A., & Mohamed, C. (2022). Towards a new method for classifying employee performance using machine learning algorithms. *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 1–5. https://doi.org/10.1109/ISCV54655.2022.9806118

A G, A. G., Su, H.-K., & Kuo, W.-K. (2023). Unleashing Potential of Employees through Artificial Intelligence. *2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 204–206. https://doi.org/10.1109/ECBIOS57802.2023.10218636

A G, A. G., Su, H.-K., & Kuo, W.-K. (2023). Unleashing Potential of Employees through Artificial Intelligence. *2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 204–206. https://doi.org/10.1109/ECBIOS57802.2023.10218636

Deepthi, P., Shaikh, M., Nargunde, A. S., Faldu, R., Goli, G., & L, N. (2024). Deep Learning-Enabled Human Resource Analytics in Predicting Employee Performance. *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 1–5. https://doi.org/10.1109/ICONSTEM60960.2024.10568716

Vijayan, N. E. (2023). Automated Feature Engineering for Predictive HR Analytics Using Cloud-Based ETL and ML Pipelines. *Journal of Artificial Intelligence, Machine Learning and Data Science*, *1*(4), 1532–1540. https://doi.org/10.51219/JAIMLD/naveen-edapurath-vijayan/344

D., S., & Therasa, C. (2024). Automated Feature Engineering for Enhanced Employee Performance Prediction: A Machine Learning Approach. *2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 1–7. https://doi.org/10.1109/UPCON62832.2024.10983315

Yadav, R., Thakur, M. and Tiwari, V. (2023) ‘Predicting employee attrition: The impact of hybrid work’, International Journal of Novel Research and Development, 8(5), pp. 1–7.

Saxena, S., Deogaonkar, A., Pais, R. and Pais, R. (2023) ‘Workplace productivity through employee sentiment analysis using machine learning’, International Journal of Professional Business Review, 8(4), pp. 1–11. https://doi.org/10.26668/businessreview/2023.v8i4.1216.

Necula, S.-C., Fotache, D. and Rieder, E. (2024) ‘Assessing the impact of artificial intelligence tools on employee productivity: Insights from a comprehensive survey analysis’, Electronics, 13(3758), pp. 1–29. https://doi.org/10.3390/electronics13183758.