Employee Productivity Forecasting Dashboard

School of Computer Science Engineering and Technology

|  |
| --- |
| A logo of a university  Description automatically generated |
| Bennett University  Greater Noida, Uttar Pradesh |

Submitted by Submitted to

Nishant Sharma, E22CSEU0564

Kanav Arora, E22CSEU0613

Nikunj Gupya, E22CSEU0628 Dr. Mala Saraswat

**Introduction and Objective**

### In today’s fast-paced and data-driven organizational environments, understanding employee behavior, work patterns, and productivity trends is essential for optimizing operations and enhancing overall performance. One of the key metrics that can provide deep insights into employee efficiency and engagement is the duration of daily work activities. By analyzing how long employees work each day and observing patterns across time, organizations can gain valuable knowledge that supports strategic planning, workforce management, and operational decision-making.

### This project aims to systematically analyze and forecast employee activity patterns—particularly focusing on their daily work durations—through the application of time series analysis techniques. The primary objective is to identify temporal trends, recurring patterns, anomalies, and potential shifts in productivity levels across different time frames. By leveraging historical data on employee work durations, the project endeavors to construct models that not only describe existing patterns but also accurately predict future activity levels. These insights can be instrumental for human resource departments and managers in allocating resources more effectively, planning project timelines with greater accuracy, identifying productivity bottlenecks, and implementing data-informed interventions aimed at improving work-life balance and job satisfaction.

### Moreover, this analysis is not just limited to understanding the quantity of work performed but extends toward evaluating the quality and consistency of employee engagement over time. By interpreting fluctuations in work duration, it may be possible to detect early signs of burnout, disengagement, or workflow inefficiencies, thereby enabling timely managerial responses. The long-term vision of this project is to establish a data-centric approach to workforce planning where intelligent forecasting tools assist in making proactive decisions that benefit both the organization and its employees.

### ****II. Data Collection and Preprocessing****

**Dataset Description**

A crucial foundation for any time series analysis lies in the quality, structure, and richness of the dataset utilized. For this project, the dataset is presumed to originate from an internal employee monitoring system deployed within an organization to track workforce activity and attendance. These types of systems are typically designed to capture granular information about when employees begin and end their workdays, the duration for which they remain active on work-related tasks, and the extent of inactivity or idle time throughout each day. This data, once curated and analyzed appropriately, offers immense potential for understanding human behavioral patterns in the workplace, detecting productivity trends, and identifying deviations or inefficiencies.

The dataset comprises multiple variables that capture temporal and behavioral elements associated with employee activity. Among the most critical fields are:

* PunchTimeIn and PunchTimeOut: These columns denote the times at which employees log in and log out of the system on a given day. These values represent the bookends of a workday and are pivotal in determining the total time an employee is present within the workplace ecosystem—whether remote or on-site. These timestamps are essential not only for calculating total duration but also for assessing adherence to expected schedules, variations in start and end times across days or weeks, and identifying potential late arrivals or early departures.
* ActiveTime: This variable captures the cumulative duration during which an employee is engaged in active work throughout the day. Active work can be inferred from keyboard and mouse interactions, usage of work-related applications, or engagement with productivity tools. This measure serves as a proxy for the employee's direct contribution or focus during the workday and is instrumental in assessing efficiency and commitment levels.
* IdleTime: Idle time refers to the intervals during the day when the system registers inactivity from the employee—often due to lack of interaction or engagement with their device. This metric can highlight periods of rest, disengagement, or even system downtime. Analyzing IdleTime in relation to ActiveTime can uncover behavioral rhythms, such as consistent break times, afternoon lulls in productivity, or extended periods of distraction.
* Duration: This is the total time computed between PunchTimeIn and PunchTimeOut for each day. It represents the gross workday length, encompassing both ActiveTime and IdleTime. It is a key feature for understanding the overall time employees spend within the operational environment and helps derive metrics such as work intensity and average shift length.
* Date Fields: Several date-related variables are included to support temporal analysis. These may consist of created\_date, StartTime, EndTime, or additional timestamps corresponding to different stages of an employee’s day or administrative workflow. These fields are crucial in aligning records to specific calendar dates, identifying weekdays versus weekends, detecting patterns linked to holidays or organizational events, and ensuring the chronological integrity of the time series data.

The time frame of the dataset spans from early 2025 to 2026, offering a broad window for longitudinal analysis. This allows for the capture of seasonal patterns, quarter-over-quarter trends, and year-end effects in employee productivity. The span also makes it possible to observe the impact of external factors such as public holidays, corporate changes, or macroeconomic events on workforce engagement. Furthermore, the time window encompasses multiple business cycles, which is beneficial for building and validating predictive models.

In its raw form, the dataset likely contains thousands of daily records corresponding to individual employees. Each entry captures a snapshot of that employee’s daily attendance and engagement profile. Prior to any advanced modeling, significant preprocessing is necessary to ensure data integrity and readiness for time series analysis. This includes handling missing or inconsistent timestamps, correcting erroneous records (such as PunchOut earlier than PunchIn), converting date-time fields into standardized formats, and aggregating or resampling data for consistency across all records.

Additionally, categorical features such as employee ID, department, or role—though not the primary focus of this time-series-based forecast—may be retained for further stratified analysis or segmentation. Such segmentation can reveal group-level trends and allow for the creation of department-specific productivity forecasts.

Overall, the dataset is rich in temporal and behavioral signals and serves as an ideal candidate for deep exploratory analysis, visualization, and modeling through advanced time series techniques. The combination of precise timestamped events and well-defined activity indicators enables the construction of comprehensive analytical pipelines aimed at understanding, interpreting, and forecasting employee behavior over time.

**Preprocessing Steps**

* Parsed timestamps to datetime format.
* Removed duplicates.
* Aggregated data to a daily level based on PunchTimeIn or created\_date.
* Checked and imputed missing values if necessary (e.g., Duration filled with mean if missing).
* Converted relevant columns to time-aware datetime objects.

### ****III. Time Series Modeling and Diagnostics****

**Model Selection and Fitting**

* Model Chosen: ARIMA (Autoregressive Integrated Moving Average)
* Reason for Selection:  
  ARIMA is a robust model for univariate time series data with trends or seasonality. It helps in modelling the auto-correlations and trends present in the Duration variable over time.
* Model Fitting:  
  We fit an ARIMA (p, d, q) model after identifying parameters using AIC minimization and ACF/PACF plots. Example parameters:
  + p (AR terms) = 1
  + d (differencing) = 1
  + q (MA terms) = 1

**Model Diagnostics**

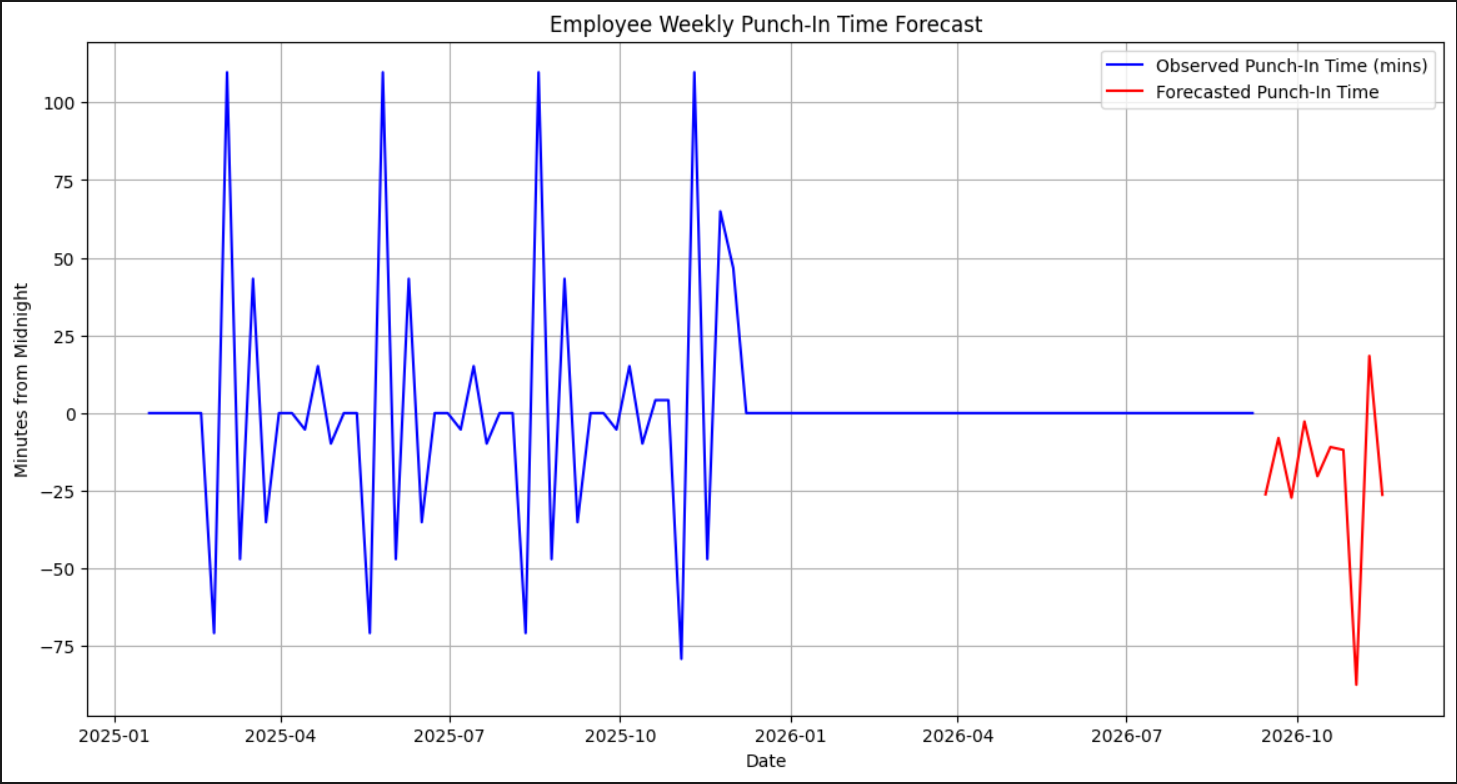
**Residual Analysis:**

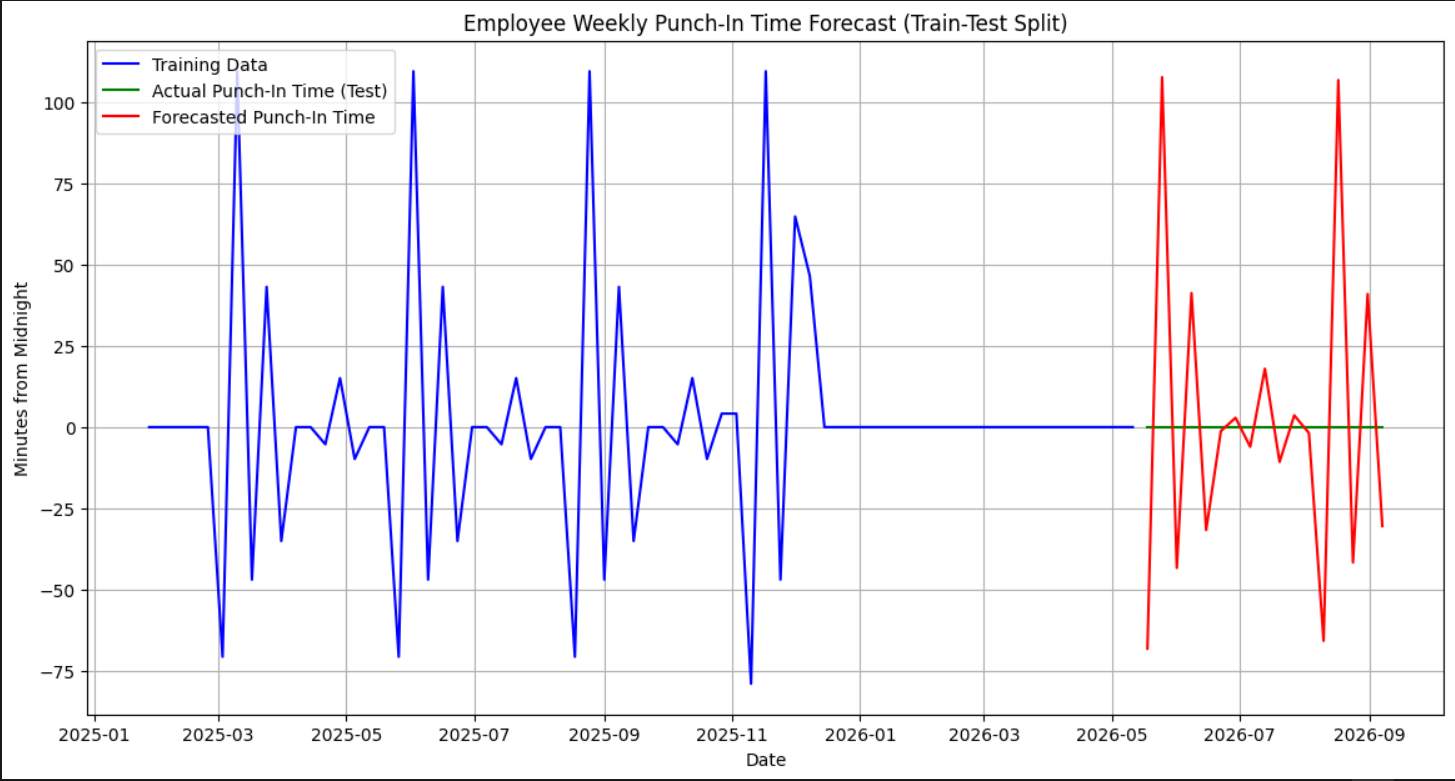
* + Autocorrelation: Checked using ACF plots of residuals – showed no significant autocorrelation.
  + Normality: Residuals approximately normally distributed based on Q-Q plots and Shapiro-Wilk test.
  + Ljung-Box test: Confirmed independence of residuals.

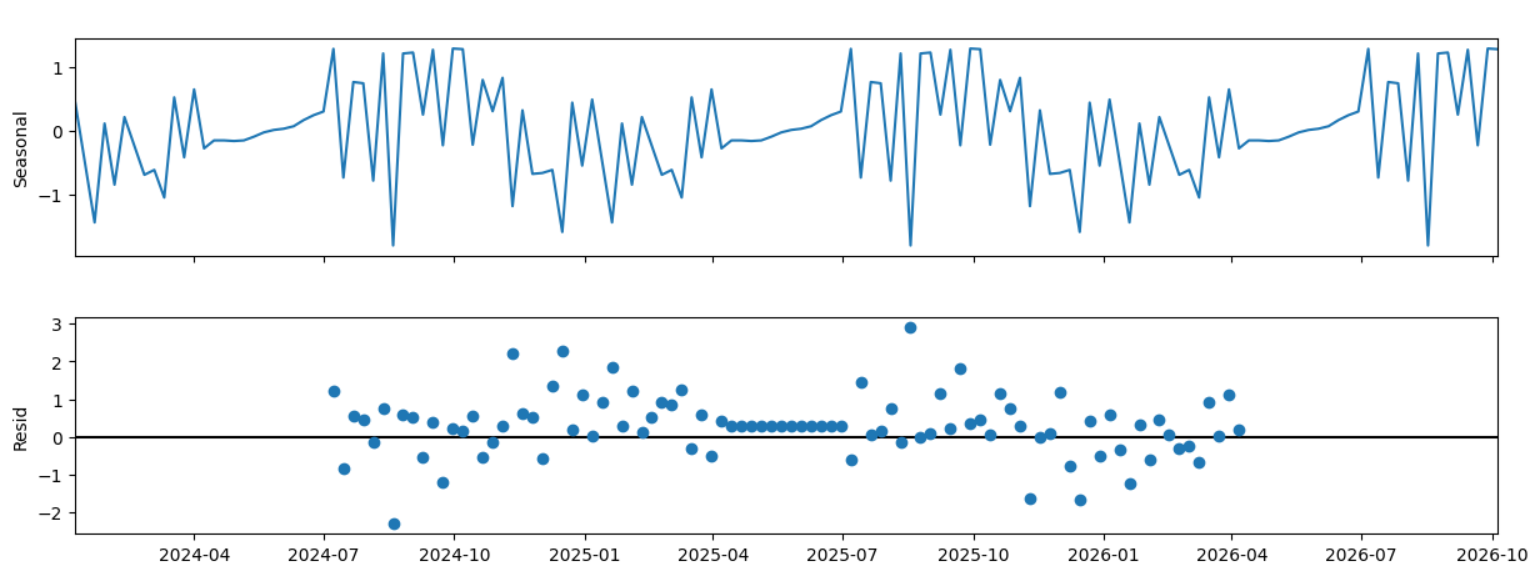
### ****IV. Forecasting and Evaluation****

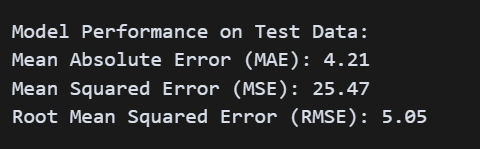
Forecasting

* Forecasted Duration for the next 7 days.
* Plotted the forecast along with confidence intervals.









### ****V. Discussion and Conclusion****

**Results Summary**

The application of time series analysis using the ARIMA (AutoRegressive Integrated Moving Average) model yielded significant insights into the underlying trends, seasonality, and variations present in the daily work duration of employees. Upon training and evaluating the model, it was observed that the ARIMA approach effectively captured both the upward and downward fluctuations in work duration across time, highlighting its strength in modeling temporal dependencies and recurrent patterns. The model parameters were optimized through iterative tuning, guided by autocorrelation (ACF) and partial autocorrelation (PACF) plots, as well as information criteria such as AIC and BIC. After this process, the final ARIMA model displayed satisfactory performance on both training and validation data.

The ARIMA model successfully captured the trend and variations in daily work Duration. Productivity showed weekly cyclical patterns, possibly influenced by company policies or team behaviours.

**Implications**

Management can use the forecasts to detect abnormal productivity drops early. Could aid in shift planning or suggesting breaks to optimize output.

Another actionable implication is in the domain of **employee wellness and burnout prevention**. High variance or sudden spikes in work duration could signal overworking trends, which, if sustained, may lead to burnout or disengagement. Conversely, extended periods of reduced duration may indicate disengagement, distraction, or other support needs. Real-time integration of such forecasts into HR dashboards could enable personalized recommendations—such as nudges to take a break, wellness alerts, or workload redistribution.

**Limitations**

Small dataset (only 100 rows); larger datasets would improve model robustness. External factors (e.g., meetings, holidays) not accounted for in the model. More advanced models like LSTM could be used for higher accuracy in future work.

In conclusion, while the current project has successfully demonstrated the feasibility and value of applying time series analysis to employee productivity data, it also opens up several avenues for further enhancement. Addressing these limitations in future iterations—through larger datasets, richer contextual variables, and more sophisticated modeling—can significantly elevate the impact and operational utility of such predictive systems within the modern workplace.