Department of Mathematics

Wearable Device Posture Forecasting using Time-series Models

by

Nitin Nandeshwar

Supervisor: Dr David Hawe

Thesis submitted in partial fulfilment of

Masters' in Data Science & Analytics

In association with

Submitted to Cork Institute of Technology

August 2020

Declaration of Authorship

I, Nitin Nadeshwar, declare that this thesis titled, 'Wearable Device Posture Forecasting using Time-Series Models ' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for the Masters' degree at Cork Institute of Technology.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at Cork Institute of Technology or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project report is entirely my own work.
- I have acknowledged all of the main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- I understand that my project documentation may be stored in the library at CIT and may be referenced by others in the future.

Signed:

Date: 24th Aug 2020

ACKNOWLEDGEMENTS

I would like to thank my supervisor, **Dr David Hawe** for his valuable inputs and contribution in this project. His guidance and motivation made it possible for me to complete the thesis successfully. I would also like to thank **Breffni Allen (CTO)** and **Ed Harnett (CEO), Habitus Posture ©** for giving me an opportunity to work on this project with them.

Abstract

Electronic wearable devices such as smartwatches and activity trackers are becoming increasingly popular for personal use in everyday life. Wearable devices are worn on or near the skin to track physiological data such as sleep, heart rate, step count and blood pressure. However, with advancements in technology, devices are now being created which can aid in posture correction. Despite companies providing both training and ergonomic office equipment for good posture, it is not always enough to stop workers from developing bad posture. The wide variety of health-related issues which stem from individuals adopting poor posture habits means that wearable technology has the potential to be immediately useful in office work environments. This thesis investigates different time series analysis method for posture like additive Holt-Winters (HW), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) and the Prophet method by Facebook. Using which changes in good posture and bad posture of average users in a business hour on weekdays can be tracked and correction for maintaining good posture can be recommended to the corporate users. The data used for time series analysis of posture is taken from Habitus Posture. To evaluate forecast accuracy as well as to compare among different models fitted to a time series, we have used the four performance measures, viz. MSE, MAE, RMSE, MAPE.

Table of Contents

List of Figures

List of Tables

1 Introduction

This project has been completed in collaboration with Habitus Posture. The company focuses on providing its clients insight on their employees' health issues like back pain. This project aims to help corporate firm (clients of Habitus) better understand their target hour of the working day where the employee's posture varies which might lead to health issues.

An early modern human has spent much of his life on work, study and other activities on his desk. Humans still focus too much on their jobs, though, that they neglect to remain sitting properly. A poor sitting posture can be detrimental in the long term, and it can have a variety of physical and mental negative side effects. Commonly known side effects include back pain, back hunch, fatigue rise, etc. Research has also shown that anyone with a slumped sitting posture is more likely than people with the correct upright sitting posture to cause negative thoughts, anxiety, frustration and sadness (Nair *et al.*, 2015). Approximately half (54%) of Americans who have low back pain sit in an unsuitable position most of their working days ('Back Pain Statistics', 2019). It is stated from the same article that 31% of men and 20% of women with back pain experience back pain that has impaired their ability to function.

Recent microelectronics advances have led to an enormous boost in research into wearable health sensing that makes small and comfortable sensors easier to wear for longer periods (weeks to months). These new devices run on low-density batteries and contain the sensors and storage space to record the person's physical movements.

According to Breffni Allen (Personal Communication), Habitus Posture is the system that allows employees to negate the effects of prolonged sitting by monitoring and coaching best posture, as well as directing users to move and exercise during office hours. This clinically evidenced approach will combat the detrimental effects of poor posture. The results will be hugely impactful - healthier employees, reduced absenteeism, and reduced visits to medical care facilities, which are an economic burden from both a medical & time loss perspective.

Posture is the way in which we stand, sit or lie down our body. Good posture allows us to stand, walk, sit and lie in positions that reduce tension on supporting muscles and ligaments when moving and bearing weight. Today's workforce has changed significantly from manual labour to physically inactive work, with many people sitting at desks or screen work for 8 hours or more a day. Studies estimate that a typical worker spends approximately 50,000 hours sitting on a chair in his lifetime (Schrempf *et al.*, 2011). Employees with an incorrect ergonomic role were correlated with these desk-based occupations, which present a health risk. Poor posture has led to a number of problems, ranging from physical health issues to psychological health impacts (Kang *et al.*, 2016). This makes the right sitting position a major concern for individuals, employers and society as a whole.

It has been documented the importance of the right ergonomic sitting position both for psychological and physical wellbeing. Despite broad acceptance that posture modifications, or ergonomic interventions such as specific hardware (e.g. adjustable chair) or other work station variations, can avoid work-related musculoskeletal disorders among computer users, there are very few reliable, objective and accurate methods to monitor posture continuously in a work environment in order to assess their success. Further, given the known financial costs of posture-related problems to both employees and organisations, it is becoming increasingly important to adopt interventions that reduce, and where possible, prevent these injuries. 'Sitting is the new smoking' coined by Dr Levine, is a popular phrase used by healthcare professionals and highlights the need to adapt the sedentary lifestyles many people find themselves living (mainpath, 2017). Wearable devices have the potential to directly contribute to individual health improvements and in turn a healthier workforce.

This thesis aims to explore the variation in good posture and bad posture of the average user during the business hour on weekdays. Besides, we are going to predict the future posture value (in minutes) spent by the average user using the forecasting method. The research questions this project aims to answer are

1. Evaluate the hour of the day when average users posture changes.
2. Compare the employees' posture as per the day of the week.
3. Forecast the future time spent by the employees in different postures.

2 Literature Review

Having poor desk posture can have a serious impact on one's physical health in ways one might not even realise. Several studies have shown that a bent, weak posture significantly decreases lung capacity, expiratory flow and lumbar lordosis compared to normal upright posture (Kang *et al.*, 2016).

(Lin *et al.*, 2006), found that sitting in a slumped position significantly decreased lung capacity and expiratory flow. This is believed to occur because slumped posture may be impeding one's diaphragm movement and adding compression to organs. A separate study by (Ghanbari *et al.*, 2008), also found a significant correlation between forward shoulder position (FSP) and respiratory values. The respiratory values they measured decreased with increasing FSP degree.

Furthermore, musculoskeletal disorders are among the most common of those physical health problems associated with poor desk posture (Van Eerd *et al.*, 2016). Musculoskeletal disorders (e.g., muscles, tendons, ligaments, nerves, dishes and the vessels of the blood), are wounds which affect the movement of the human body. According to the American Chiropractic Association (2020), 95% of all headaches are known as primary headaches and are associated with muscle tension in the neck. Sedentary activities such as sitting at a desk for hours increase joint irritation and muscle tension in the neck, upper back and scalp. Future head posture (FHP) is an earlier positioning of the head with respect to the sagittal plane's line of gravity. This symptom is commonly found in people who operate on a computer for an extended period of time and is closely related to a form of cervical headache induced by the poor posture of the cervical spine. (Nobari *et al.*, 2017).

Sitting with incorrect posture can lead to a variety of health-related issues for desk-based workers. Musculoskeletal disorders and headaches can have a hugely negative impact on an individual's ability to work and results in significant absenteeism from work. (Van Eerd *et al.*, 2016) indicated that the main causes of work-related injuries in Canada and the United States were the higher extremity and low back pain. Around 2002 and 2013, in Ireland, 50 per cent of self-reported working-related illnesses were attributed to musculoskeletal disorders. However on an average number of days absents was greater than the average of all other illness except anxiety, stress and depression. (Russell *et al.*, 2018). Many workers often still attend work despite feeling that they should have taken sick leave, this is a phenomenon known as sickness presenteeism (Aronsson, Gustafsson and Dallner, 2000). One study which included 654 computer workers from across five different companies, reported that for most computer workers with neck, shoulder, hand or arm symptoms, there was more productivity lost as a result of decreased performance at work than from physical absence at work (van den Heuvel *et al.*, 2007). Both absenteeism and individuals with reduced productivity due to functional limitations are not only health concerns but are also a financial concern as it costs individuals and businesses substantial amounts of money each year (Epstein *et al.*, 2012).

(Womersley and May, 2006) examined the source of the backaches of the sitting people: Was this due to a poor posture and a long sit? They divided the subjects into two classes of people with back pain and others and videotape their pose. In the postural backaches, they find that continued sitting times and continuous study length are considerably longer and that the degree of flexion is also markedly higher in relaxed sitting than in the other class.

(Goossens, Netten and Van der Doelen, 2012) had analysed the influence of smart chair on sitting habits of office workers in a field test for 4 weeks. For this test, 40 office workers (13 male, 27 female) were picked. They were divided into three groups. A monitoring group, a group receiving a sitting instruction and a group receiving sitting instruction and suggestions each hour they sit on their posture. The findings indicate that both the group that received guidance and the group that received input have an effect on the average change in the basic posture. Over time, the effect decreases. The control group was not affected

Smart Cushion is a textile sensing device that reliably and non-invasively analyzes human sitting positions (Xu *et al.*, 2013).In this paper, he used the electrical sensor for sensing and its electrical properties, such as offset, scaling, crosstalk and rotation. Many effective techniques have been proposed to improve recognition levels of sitting positions, including sensor calibration, data representation and dynamic time warping classification.

(Zhang *et al.*, 2014) proposed the use of the Microsoft Kinact sensor for a new approach for the recognition of human posture that could recognize any position defined by users automatically. In order to produce 9 features representing certain body parts such as the forearms, the thigh, etc, the skeletal information derived from the user's profound image was used. Such features are fed into SVM to create machine-learning models for posture that then are used for recognizing predefined positions. Totally 22 different positions including body, arm, leg positions and PCA analyzes were collected and showed that the feature was generally well separated in space. Besides,k-fold cross validation used with k =10 for increasing the calculation efficiency and the final cumulative accuracy of 99.14 per cent obtained.

(Wang *et al.*, 2016) combined the multi-image processing with image depth obtained from the Kinect sensor to identify the 5 distinctive human sittings, standing, stooping, kneeling and lying positions successfully. The positive detection rate of the test information is greater than 97 per cent, although the test subjects may not be confronted by the Kinect sensor and may have different statures, using the neural net of pre-trained LVQ (learning vector quantizations) details.

A wearable posture corrector that recognizes the good posture and shows whether the pose is bad has been discussed. (Bramhapurikar *et al.*, 2018) .The correct position of the posture is calculated using the Flex sensor bending feature. He used a vibration motor to detect bad posture on this unit, which serves as a warning signal to the user. He used Bluetooth to get a message on the smartphone that indicates bad posture and prevent the use of motors., which decreases the battery usage To make it more portable, the system was tested with the Nokia 5110 smartphone The device is initially calibrated and the threshold for the correctness of the posture is then established. It takes only a few seconds to complete the calibration. When the posture is slumped it shows "bad posture" on the smartphone and also activates the vibration motor to alert the user. This rapid reaction is important to warn the consumer instantly that it is in bad shape so that it can correct itself. The message is sent by Bluetooth module HC-05 to Smartphone. The machine begins to evaluate again whether the posture is a poor pose or good pose, after returning to the right position. When the user will not get back indefinitely then the error signal and message will be returned for the posture correction.

As per the literature study, it is seen that good posture and bad posture research have never centred on other writers, primarily on a time series analysis of office employees. Their research focuses on general positions such as stand, move, sit and squat. Although some earlier research focuses on the pose, they use a wearable postural corrector that directly discusses the identification of a proper and incorrect sitting stance that helps them correct their posture, but you can't always sit in good posture. Therefore, there is a need to study the change in posture during the business hour which will, in turn, help employers to properly plan rehabilitation program during the office hour for their employees to keep them fit and avoid future related health issues

3 Methodology

This is a revolutionary shift to switch from computer simulation to the estimation of time series. It is a challenging yet enriching experience that has helped to better understand how machine learning can be applied to posture forecasts. The statistical model aimed at estimating the value of an unknown variable. A time series has the time (t) as an independent variable and an output variable (y). The response of the model is for y at t, y_t' is the expected value.

3.1 Data Preparation

3.1.1 Data description

1.Habitus User Database Dataset:

- user id: Unique identification for each employee (user).
- created date: date & time of sensor data for each user
- Modified date: Modified data of created date
- Corporate code: Unique identification code for corporate user
- Role id: Type of role performed when using habitus product
    - User = 1
    - Administrator = 2
    - CorporateAdministrator = 4
- Gender id: Gender label of each user
    - Nonspecified = 0
    - Male = 1
    - Female = 2
- Height: Height of users in centimetre (cm)
- Weight: Weight of users in kilogram (kg)
- Age Range id: Categorical label for different range of age
  Below Figure 3.1Figure 3.1 shows the user age distribution.

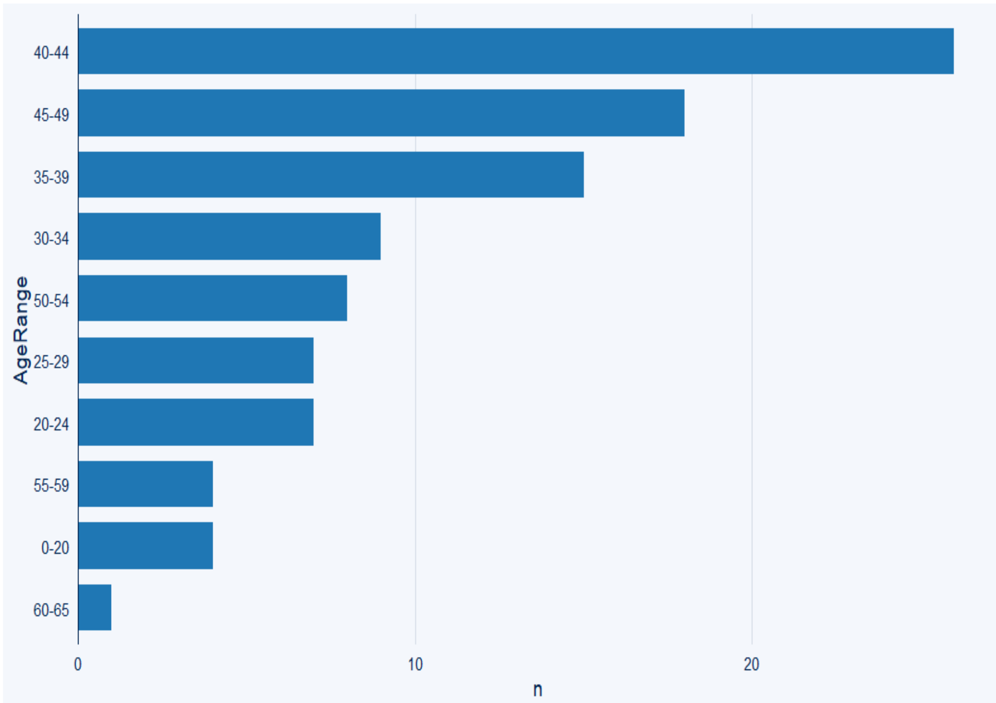

Figure 3.1 Bar plot of Age range of employees

- Job Type id: Type of job each user
  - Low = 0
  - Medium = 1
  - High = 2

2.Aggregate by Day Dataset:

- Id: Unique identification for each observation.
- Date: Date on which data is collected
- GoodPostureTimeSeconds: Time(seconds) spent when user posture was in good sitting position
- BadPostureTimeSeconds : Time(seconds) spent when user posture was in bad sitting position
- UnknownPostureTimeSeconds: Time spent by user nether in good & bad position
- UserId : Unique User identification number.

3.Aggregate by Status Change Dataset:

- Id: Unique identification for each observation
- SensorId: Unique identification for each sensor.
- Posture: Position of the top of the body including chest, shoulder, neck and head. Here posture label in which position user is like Unknown = 0, Bad = 1, Good = 2 .
- DateFrom: User recorded start time as per posture.
- DateTo: User recorded end time as per posture.

How the data is collected?

According to Breffni Allen (Personal Communication), Pilot data is collected from a large employer in Munster Here 80 participants were given Habitus Posture device to wear every workday for 5 weeks. The Habitus Posture device works when user adopting to bad posture is determined by an inbuilt 6-Axis Accelerometer/Gyroscope sensor and a notification is sent to the haptic motor for it to vibrate and alert the user

A mobile app compatible with iPhone and Android phones is the key component as it will advise on getting a good posture, communicate in real-time with the sensor, advise on how to achieve good posture, trigger the notification to the user, present a view of the current and past posture time and send anonymised data to the portals. Through the portal, the posture data is collected and saved in Aggegrated by status change dataset and Aggegrated by days dataset. All the mandatory personal details are filled by the user on signup which is collected and saved in Habitus user database dataset through the portal.

3.2 Final Datasets

As Aggregate by days dataset contains posture data in day format while Aggregate by status change dataset contains posture data in time second format. Based on the problem statement we have to analyzed posture change during business hour. So by using Aggregate by status change dataset, posture data is transformed from time second to hour and a time series format for good posture and bad posture is developed. Here data preparation is done in Rstudio using R programming. Libraries used are dplyr, tidyr , VIM and ggplot2.

Following are Features of the final dataset:

- Dates: Filtered dates from 27th Jan to 28th Feb
- Hours: Working Hours between 8 AM to 6 PM
- GP_Sec: Total users Good Posture time in seconds per hour
- BP_Sec: Total users Bad Posture time in seconds per hour
- Active_User: Number of User active per hour

3.3 Time Series Analysis

3.3.1 Basic of Time Series.

Every sequence of sequential data points is a time series. There are two key elements in each time series: time and value allocated to the correct time point.

Generally for two main purposes, a time series analysis is done:

- Understand the way the method operates by analyzing the past records can model and identify the key parameters which affect the time series and its components are known.
- Predicting the future values of the time series, using the appropriate model that has been trained on historical data of time series.

Time Series Components

The key parameters that are responsible for bringing about changes in a time series, also known as the components of time series, are shown in Figure 3.2

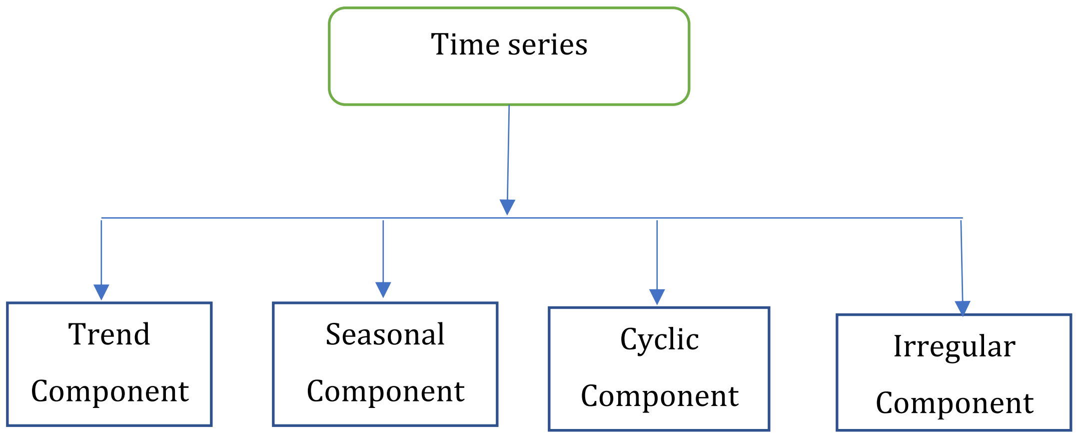

Figure 3.2 Time Series Components

These components can be combined in two ways. Generally, they are considered as Additive or Multiplicative (Rob J Hyndman, 2013a)

For Additive

$$y(t) = T(t) + S(t) + C(t) + \varepsilon(t) \qquad 3.1$$

For Multiplicative

$$y(t) = T(t) \times S(t) \times C(t) \times \varepsilon(t) \qquad 3.2$$

In equation (3.1 and 3.2) $y(t)$ represents the measure captured at time step t

Trend component $T(t)$**:** The trend is the time series' long-term pattern. A trend may be positive or negative depending on whether the time series shows a rising pattern or a declining pattern over the long span.

Seasonal component $S(t)$**:** Seasonality is when the time series periodically repeat at a constant frequency.

Cycles component $C(t)$**:** Any pattern that shows ascending and descending movements around a specific trend is a cyclical pattern.

Irregular Component $\varepsilon(t)$: This component is not predictable and it takes any random value. It is also known as residuals.

3.3.2 Processing Data for Time-series Analysis

a. Stationarity Test

Before selecting any time series model, data should follow the assumption of models To order to do so, the stationary test must be completed. Stationary data does not change the variance and auto-correlation over time (Brownlee, 2016).

Augmented Dicky-fuller test

The Dicky-Fuller test is the most popular method to check data to be stationarity. It is used to find the root of the time-series data to check whether the data set has a unit root. (Brownlee, 2016) The null and alternative hypothesis of this test is:

Null Hypothesis: The time series has a unit root (value of a =1) and it is stationary.

Alternate Hypothesis: The time series has no unit root and it is non-stationary.

If data is non-stationary then it will reject the null hypothesis and the series can be linear or difference stationary.

Test for stationarity: If the test statistic (p) is less than the critical value, data is stationary and has a unit root. When p is greater than the critical value, data is non-stationary. The critical value for this test is 0.05(Brownlee, 2016).

b. Making Data stationery

After performing the stationary test, if data is stationary then it is alright to go for the forecasting algorithms otherwise one has to make data stationery. To make data stationary first one have to visualize the real data by plotting and try to guess the properties of the data such as linear trend, seasonality, random walk and so on.

Differencing

By differencing one can eliminate trend and seasonality which stabilize the mean of a time series. Also, by visualizing the ACF plot of the data, one can identify non-stationarity. For a stationary time series, the ACF will sharply reduce to zero, while the ACF of non-stationary time series reduces slowly(Rob J Hyndman, 2013b).

Seasonal differencing

A seasonal difference is a difference between a present value and the past value from the same season. The equation

$$y_t' = y_t - y_{t-m} \qquad 3.3$$

In the above equation (3.3), “m” is the number of seasons and it is also known as “Lag-m differences”.Seasonal differencing works well if the time series data has strong seasonality component otherwise it makes no difference after first differencing (Rob J Hyndman, 2013b).

3.4 Forecasting Algorithms

3.4.1 Holt-Winters Exponential Smoothing method

Holt-Winters Exponential Smoothing (HW) method is one of the simplest and most widely used methods in the industry for data with seasonal patterns and trends. It comprises of the forecast equation and three smoothing equations (3.5,3.6,3.7): one for the level ℓ_t, one of a trend component b_t, and one of a seasonal index s_t, each with a corresponding smoothing parameters: α, β, γ, respectively(Gardner, 1985).

The two main variations of the HW method differ in the nature of the seasonal equation. One method uses an additive seasonal equation (called additive HW) and the other a multiplicative seasonal equation (called multiplicative HW). The additive method is preferred, when the seasonal variations are of the same magnitude throughout the data set, while the multiplicative method is preferred when the magnitude of seasonal variations increases with time. From Figure 4.6 the posture amplitude is nearly constant throughout the time so the additive seasonal equation is preferred and below equations (3.4 to 3.7)(*Rob J Hyndman*, 2010).

Suppose the time-series, y, where y = y1, y2,… y_t for each time period t, m is the seasonal period (e.g., m = 10 for a business hour from 8 am to 6 pm). Let $\hat{y}_{t+h|t}$ represent the forecast of the time-series value h-steps ahead of the observed value at time t.

The basic equations for additive HW method (Gardner, 1985) are as follows:

$$Forecast: \hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \qquad 3.4$$

$$Level:\ \ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \qquad 3.5$$

$$Trend: b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \qquad 3.6$$

$$Seasonal:\ s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \qquad 3.7$$

Above, equation (3.5) l_t is the level of posture at time t and it is a weighted average of the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(\ell_{t-1} + b_{t-1})$,equation (3.6) b_t is the trend of posture at time t and it is a weighted average of the estimated posture trend at time t based on $(\ell_t - \ell_{t-1})$ and b_{t-1}, the previous estimate of the trend., equation (3.7) s_t is the seasonal component at time point t and it is a weighted average between the current seasonal index, $(y_t - \ell_{t-1} - b_{t-1})$ and the seasonal index of the same season last year s_{t-m} The parameters estimated in the HW method, α, β, and γ are restricted to lie between 0 and 1.

3.4.2 ARIMA and its variants

3.4.2.1 AR(p)

Assumption: The time series is stationary

An autoregressive model is when a value y_t from a time series is regressed on previous values y_{t-p} from that same time series. The autoregressive model is expressed in equation (3.8)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \qquad 3.8$$

The above equation can be denoted in terms of summation in equation (3.9)

$$y_t = c + \sum_{p=1}^{m} \phi_p y_{t-p} \qquad 3.9$$

Here p is the order of autoregressive model Generally, AR(p) is a pth-order autoregression, is a multiple linear regression in which the value of the series at any time t is a (linear) function of the values at times t−1,t−2,…,t−p (Dr Iain Pardoe, 2018)

The coefficient of correlation between two values in a time series is called the autocorrelation function (ACF). It is denoted as

$$\mathrm{Corr}\left(y_t, y_{t-p}\right) \qquad 3.10$$

When p=1 in equation (3.10), it is lag 1 autocorrelation i.e correlation between values that are one time period apart. Generally, lag p autocorrelation is the correlation value that is p time apart.

For the AR(p) model, it only measures the correlation between y_t and $y_{t-p}$and filter out the linear influence of the random variables that lie in between (i.e., $y_{t-1}, y_{t-2}, \ldots, y_{t-(p-1)}$), which requires a transformation on the time series. This transformation can be done using the partial autocorrelation function (PACF). So PACF is most useful in identifying p-value(Dr. Iain Pardoe, 2018).

3.4.2.2 MA(q)

Assumption: The time series is stationary

Instead of using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model as shown in equation (3.11).where ε_t is white noise and MA(q) model, moving average of order q (Rob J Hyndman, 2013c).

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \qquad 3.11$$

Above equation (3.11) can be rewritten in terms of summation as

$$y_t = \mu + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} \qquad 3.12$$

$$\mu = \frac{1}{q}\sum_{t=1}^{q} y_{t-i} \qquad 3.13$$

As described in the equations (3.11,3.12,3.13) the MA component is the linear combination of the latest q residuals with $\theta_{i\in[1..q]}$the weight of each error correction. The component $\varepsilon_{t-i}, i \in [1, q]$ is called a random shock which is assumed to follow a normal distribution with zero mean and a constant variance σ^2.

3.4.2.3 ARMA(p,q)

Assumption: The time series is stationary

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad 3.14$$

The ARMA(p,q) model is the combination of the previously mentioned AR(p) and MA(q), mathematically the model is simply the sum of these models as described by equation (3.14). But the weakness of these models is that they assume that the time series that we are working on is stationary which is not the case in most real-life cases. Thus, one needs to make sure that any time series modelled with this function needs to be stationary, for this purpose, one needs to use the differencing operation to stationarize the series. This operation is handled by the ARIMA model which generalizes the ARMA model for non-stationary time series (Stephanie, 2019).

Backshift Operator

Generally backshift operator an operator which changes an observation to the one from the previous time step (Nau, 2014). It is denoted by B and its equation for time series Y_t is

$$BY_t = Y_{t-1} \qquad 3.15$$

In general, for any k integer equation (3.15) can be rewritten as

$$B^k Y_t = Y_{t-k} \qquad 3.16$$

3.4.2.4 ARIMA(p,d,q)

Equation (3.14) can be rewritten as

$$y_t - \sum_{i=1}^{p} \phi_i y_{t-i} = \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad 3.17$$

The left-hand side of equation (3.17) represents the autoregressive part while the right-hand side represents a moving average part. So backshift operator of equation (3.16), equation (3.17) can be rewritten as

$$\left(1 - \sum_{k=1}^{p} \phi_k B^k\right) y_t = \left(1 - \sum_{k=1}^{q} \theta_k B^k\right) \varepsilon_t \qquad 3.18$$

where B is the backshift operator; p; q ≥ 0

A way to generalize this model a bit is to consider the ARIMA(p,d,q) model, where the 'I' stands for integrated. In backshift notation, it looks like this

$$(1-\sum_{k=1}^{p}\phi_k B^k)(1-B)^d y_t = (1-\sum_{k=1}^{q}\theta_k B^k)\varepsilon_t \qquad 3.19$$

the factor (1 − B) on the left-hand side transforms the data by letting

$Z_t = y_t - y_{t-1}$, generating a new time series Z_t. This is referred to as taking the first difference of the series, and the d means that this is done d times. The model is therefore called 'integrated' because the model is still written in terms of y_t, with the differencing being a part of the model itself.

Let us simplify our notation, by introducing the following naming conventions (Box, 2015)

$\nabla := (1-L)$	the (first) difference operator
$a_p(B) := (1-\sum_{k=1}^{p}\phi_k B^k)$	the AR(p) backshift polynomial
$m_q(B) := \left(1-\sum_{k=1}^{q}\theta_k B^k\right)$	the MA(q) backshift polynomial

With these, we can write our ARIMA(p,d,q) model as

$$a_p(B)\nabla^d y_t = m_q(B)\varepsilon_t \qquad 3.20$$

3.4.3 SARIMA Model

In many time series analysis cases, there is a seasonal pattern. For example, in case of posture daily data, there could be a daily pattern (all business hour are same) or weekly pattern (all Friday are similar, etc) For such time series we might consider the following model.

$$(1-\sum_{k=1}^{P}\Phi_k (B^s)^k)(1-B^s)^D y_t = (1-\sum_{k=1}^{Q}\theta_k (B^s)^k)e_t \qquad 3.21$$

This equation (3.25) is somewhat similar to equation (3.19) of ARIMA model, but where all lags are multiples of some positive integer s called the seasonality. For daily patterns in posture daily data,s =10.One can name the parts of this model similar to the regular ARIMA by letting

$$\nabla_s := (1 - B^s)$$
$$A_P(B^s) := (1 - \sum_{k=1}^{P} \Phi_k (B^s)^k)$$
$$M_Q(B^s) := (1 - \sum_{k=1}^{Q} \theta_k (B^s)^k)$$

By substituting above equations in equation (3.25), we get below equation (3.22)

$$A_P(B^s)\nabla_s^D y_t = M_Q(B^s)e_t \qquad 3.22$$

Now, there is no reason to expect that the series e_t is uncorrelated, so a natural step is to continue by fitting a regular ARIMA model to the innovations

$$a_p(B)\nabla^d e_t = m_q(B)\varepsilon_t \qquad 3.23$$

where ε_t is (hopefully) an iid series. Substituting for ε_t in equation (3.22) and (3.27), we can write the so-called Seasonal ARIMA model as

$$a_p(B)A_P(B^s)\nabla^d\nabla_s^D y_t = m_q(B)M_Q(B^s)\varepsilon_t \qquad 3.24$$

and refer to it as the ARIMA(p,d,q)(P, D, Q)$_s$ model, which contains all necessary information-namely the degrees of all polynomials, both degrees of differencing, and the seasonality

3.4.4 Prophet method

The prophet is an open-source forecasting tool published by Facebook's core data science team and is available in Python (van Rossum, 1995) and R. Prophet is developed for typical Facebook issues such as predicting user activities. This makes the Prophet method convenient for predicting seasonalities, special events, data with holidays, data showing outliers and data with the varying trend.

The Prophet method uses a framework called “Analyst-in-the-Loop" as shown in Figure 3.3 below.

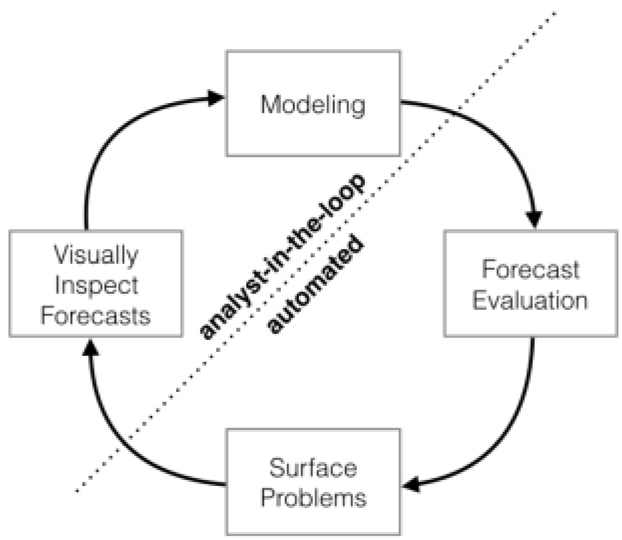

Figure 3.3: Analyst-in-the-Loop Modeling reproduced from (Taylor and Letham, 2018)

The framework is double-sided where on one side model fitting is automated assuming that the user has no statistical knowledge, while on other side the framework allows the same user to input information based on their domain/industry knowledge.

The Prophet procedure is an additive regression method which belongs to the Generalized Additive Model (GAM) family with the following components and functional form:

$$y(t) = b(t) + s(t) + h(t) + \epsilon_t \qquad 3.25$$

In equation (3.25), $b(t)$ captures trend in the time-series, s(t) captures time-series seasonality, $h(t)$captures holidays or special events in the time-series and ϵ_t is an irreducible error term. In any instance of the Prophet method, only ϵ_t is always present, the remaining three terms may not always be present, as they have to be provided by the user.

Trend Model

The Prophet library implements a piece-wise linear trend model where the growth rate remains constant and nonlinear trend model where growth rate decreases with time t(Taylor and Letham, 2018).

$$b(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \qquad 3.26$$

Here k is the growth rate, δ is the rate adjustments, m is the offset parameter, and γ_j is set to $-s_j \delta_j$to make the function continuous

The growth rate is allowed to change by defining changepoints. Suppose there are S changepoints at times s_j , $j = 1, \dots, S$.By defining a vector of rate adjustments $\delta \in R^S$, then δ_j is the change in the rate that occurs at the time s_j. The rate at any time t is then the base rate k, plus all of the adjustments up to that point: $k + \sum_{j:t>s_j} \delta_j$.By defining a vector $a(t) \in \{0,1\}^s$ then

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j \\ 0, & \text{otherwise} \end{cases}$$

The rate at time t is then $k + a(t)^T\delta$. As the rate k is adjusted, so the offset parameter m is also been adjusted to connect the endpoints of the segments. The correct adjustment at changepoint j is easily computed as

$$\gamma_j = (s_j - m - \sum_{l<j} \gamma_l)(1 - \frac{k + \sum_{l<j} \delta_l}{k + \sum_{l\leq j} \delta_l})$$

Seasonality

As Seasonality represents periodic changes (daily/weekly/monthly/yearly seasonality) in the time-series By using Fourier series in seasonality model it increases the model flexibility(Taylor and Letham, 2018).

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right) \qquad 3.27$$

The Prophet method uses a curve fitting technique for the time-series fit. The method is fitted automatically using Stan code (Carpenter *et al.*, 2017) that takes seasonality, trends, and holidays into account. Prophet's robustness, ease of configuration and being fast to fit attracts non-experts and users with limited statistical knowledge to deploy Prophet within their organization.

One weakness of prophet is that it cannot handle time series without timestamps as it uses dates to infer frequency and build the model, this weakness is dealt with in our code by returning the mean value of the train data set as the forecast.

4 Results

4.1 Exploratory Data Analysis

4.1.1 Good Posture Visualization:

Good Posture Scatter Plot:

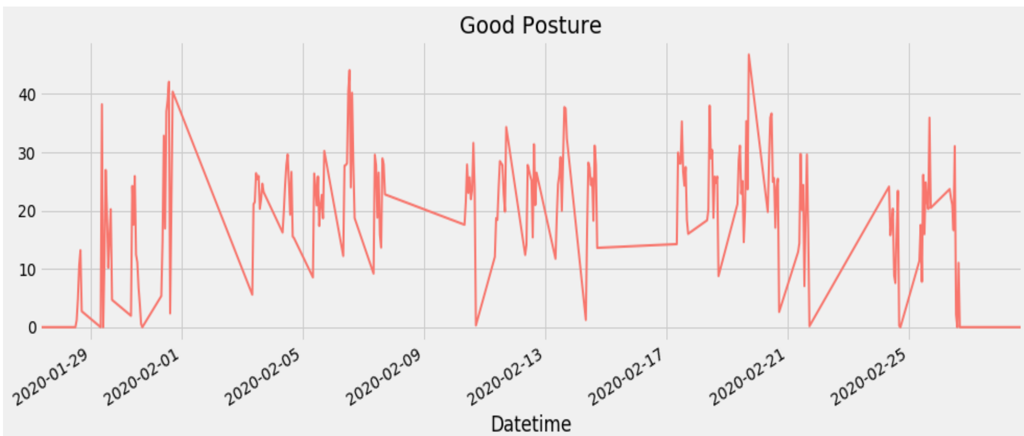

Figure 4.1: Good Posture in minutes Vs Datetime

From **Error! Reference source not found.** good posture time series scatter plot, there a lot of null values in the first week and fifth week due to missing values, which creates noise in data. While from the second week to the fourth week there is not much change in trend so data might be stationary.

Good Posture Boxplot:

Below **Error! Reference source not found.** shows the time spent by the average user in good posture as per weekday. As we can see nearly all good posture distribution on a weekday are skewed except on Tuesday as mean is nearly same as the median. On Wednesday there is an observable outlier of value near to 60 minutes. Time spent by the average user in good posture is lowest on Friday as per the median value.

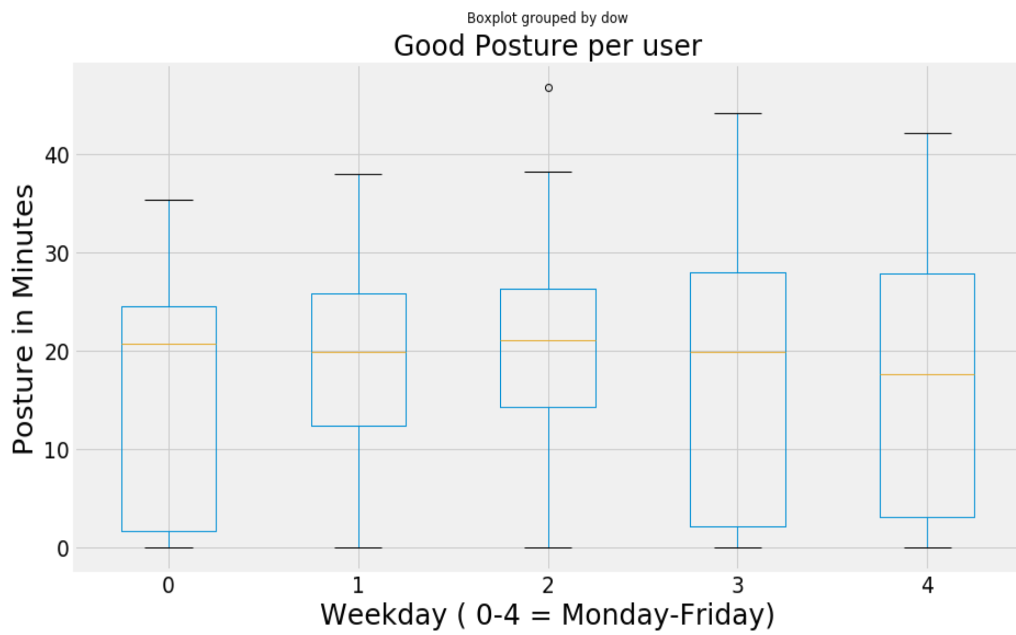

Figure 4.2: Good posture in minutes Vs day of the week (dow)

4.1.2 Bad Posture Visualization:

Bad Posture Scatter Plot:

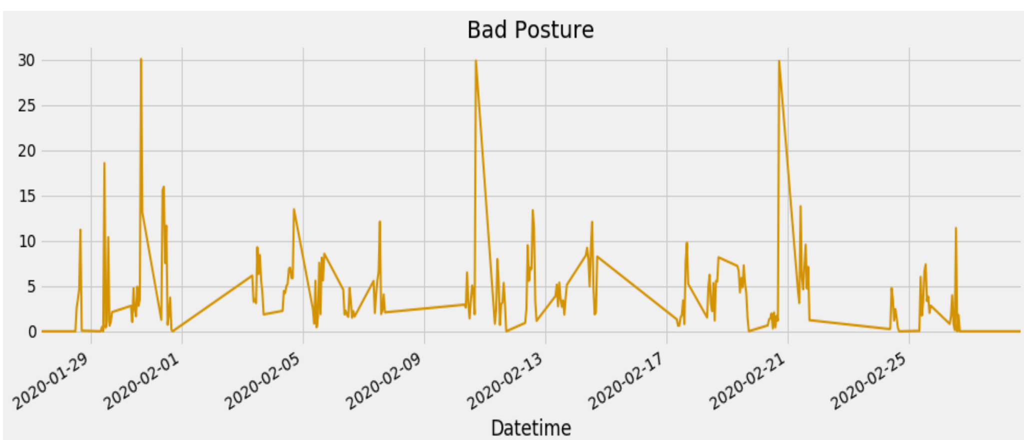

Figure 4.3: Bad Posture in minutes Vs Datetime

From Figure 4.3 bad posture time series scatter plot, first and the fifth week of data is noisy as there are a lot of null values due to missing data. While there might be a decreasing trend of small value from the second week to the fourth week. On the third and fourth week, there is a high peak due to time spent after 4 pm in bad posture is high on Monday of the third week and Thursday of the fourth week.

Bad Posture Boxplot:

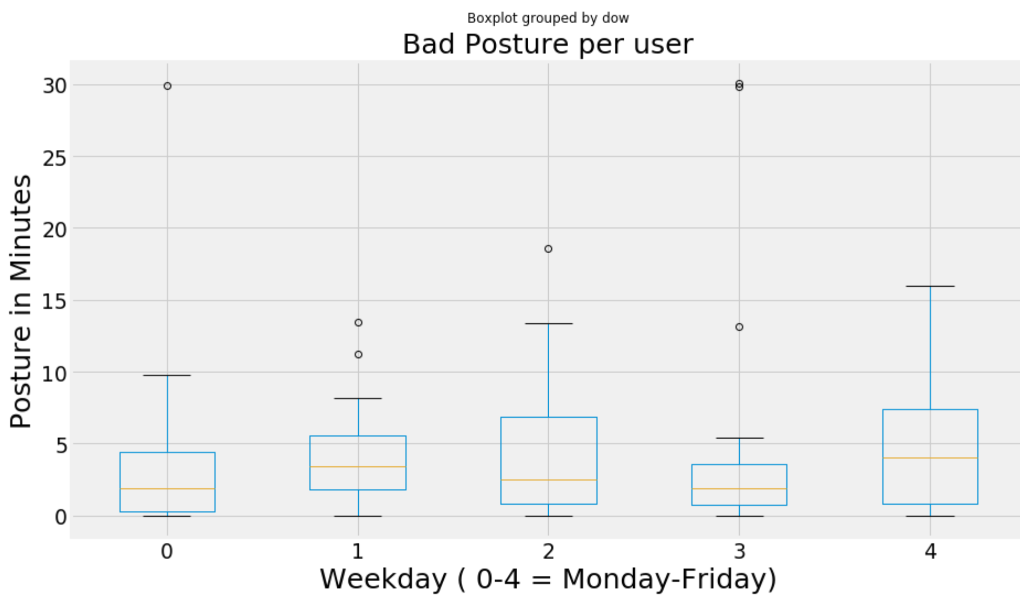

Figure 4.4: Bad posture in Minutes Vs Day of the week (dow)

Figure 4.4 shows the time spent by the average user in bad posture as per weekday. As we can see that all the bad posture distribution on a weekday are skewed and there is an observable outlier on Monday, Tuesday and Thursday. The time spent by average users in bad posture in minimum on Thursday while maximum on Friday.

4.1.3 Active Users Visualization:

Active User Scatter Plot:

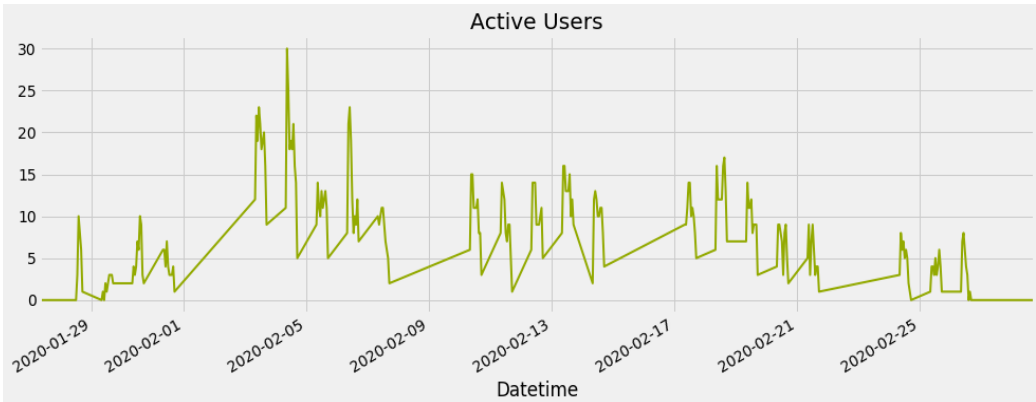

Figure 4.5: Number of active users Vs Datetime

From the above Figure 4.5, Active user time series scatter plot, the first week of data is noisy while there is a decreasing trend of a number of an active user from second to the fifth week.

4.1.4 A difference between Good & Bad Posture Visualization

Difference Posture Scatterplot:

Instead of analyzing both good posture and bad posture time series, we can analyze the difference of good posture and bad posture which will be easy to understand. The unbalanced distribution of time spent in good posture and bad posture will not impact our model as suggested by the industry partner.

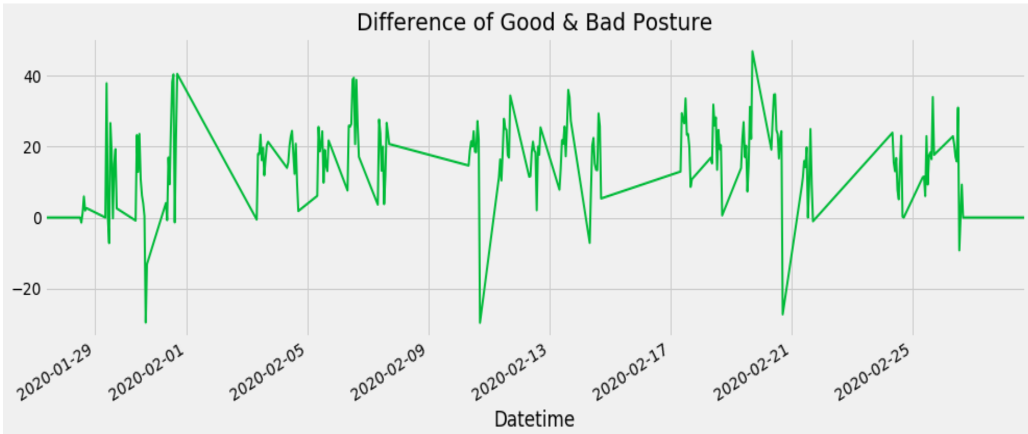

Figure 4.6: Difference Posture in minutes Vs Datetime

From Figure 4.6, the first week and fifth week contain noisy data due to missing values. As from the second week to the fourth week have nearly no trend and data might be stationary which can be verified using adfuller stationary test. So if we train the model from the second week to the fourth week we will able less residual error.

Difference Posture Boxplot:

Below Figure 4.7shows the time spent by the average user in posture as per weekday. Nearly all posture distribution on a weekday is skewed as mean and median is not equal for any boxplot. There is observable two outlier on Wednesday, negative value outlier is due to bad posture while extreme positive value outlier is due to good posture. Nearly all posture median value is in the range from 15 to 20 minutes except Friday have the lowest value to around 12 minutes. If we compare skewness of difference posture distribution is less with respect to good and bad posture.

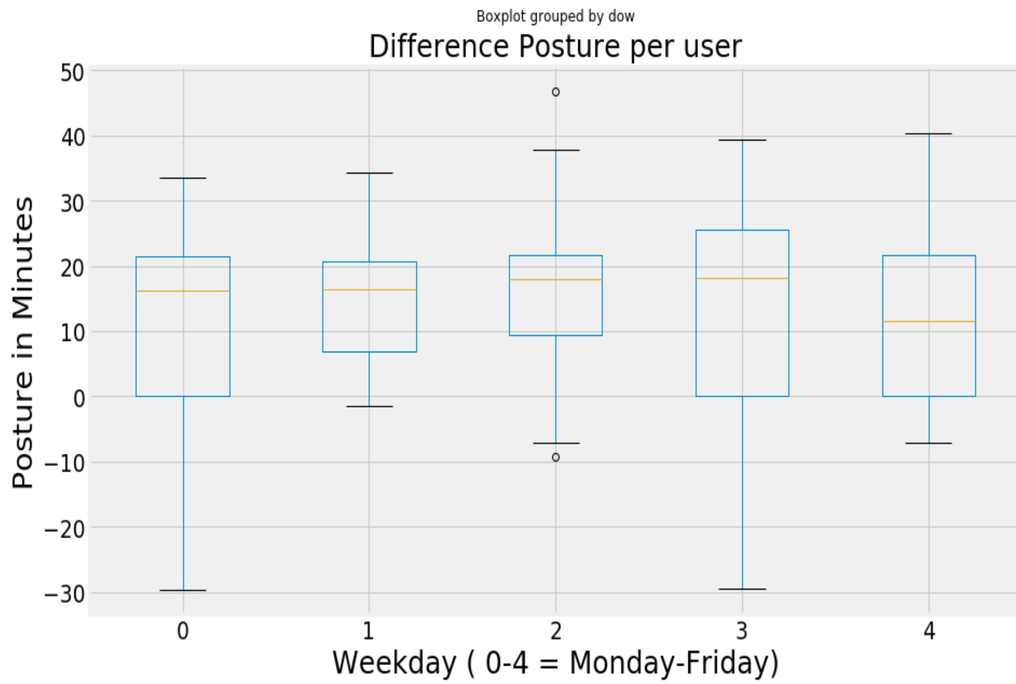

Figure 4.7: Difference posture in Minutes Vs Day of the week (dow)

Difference Posture violin plot:

Below Figure 4.8shows the difference of good and bad posture of an average user as per the business hour from 8 am to 6 pm on each weekday.

- There is a decrease in posture from 10 am to 11 am on all day of the week (dow) except Friday.
- During 1 pm to 2 pm, there is a decrease in posture every day of the week. While on Wednesday and Thursday it decreases to 3 pm.
- On Monday, Tuesday and Friday the posture gradually decreases from 3 pm to 6 pm.
- On Thursday the posture decreases from 4 pm to 6 pm.

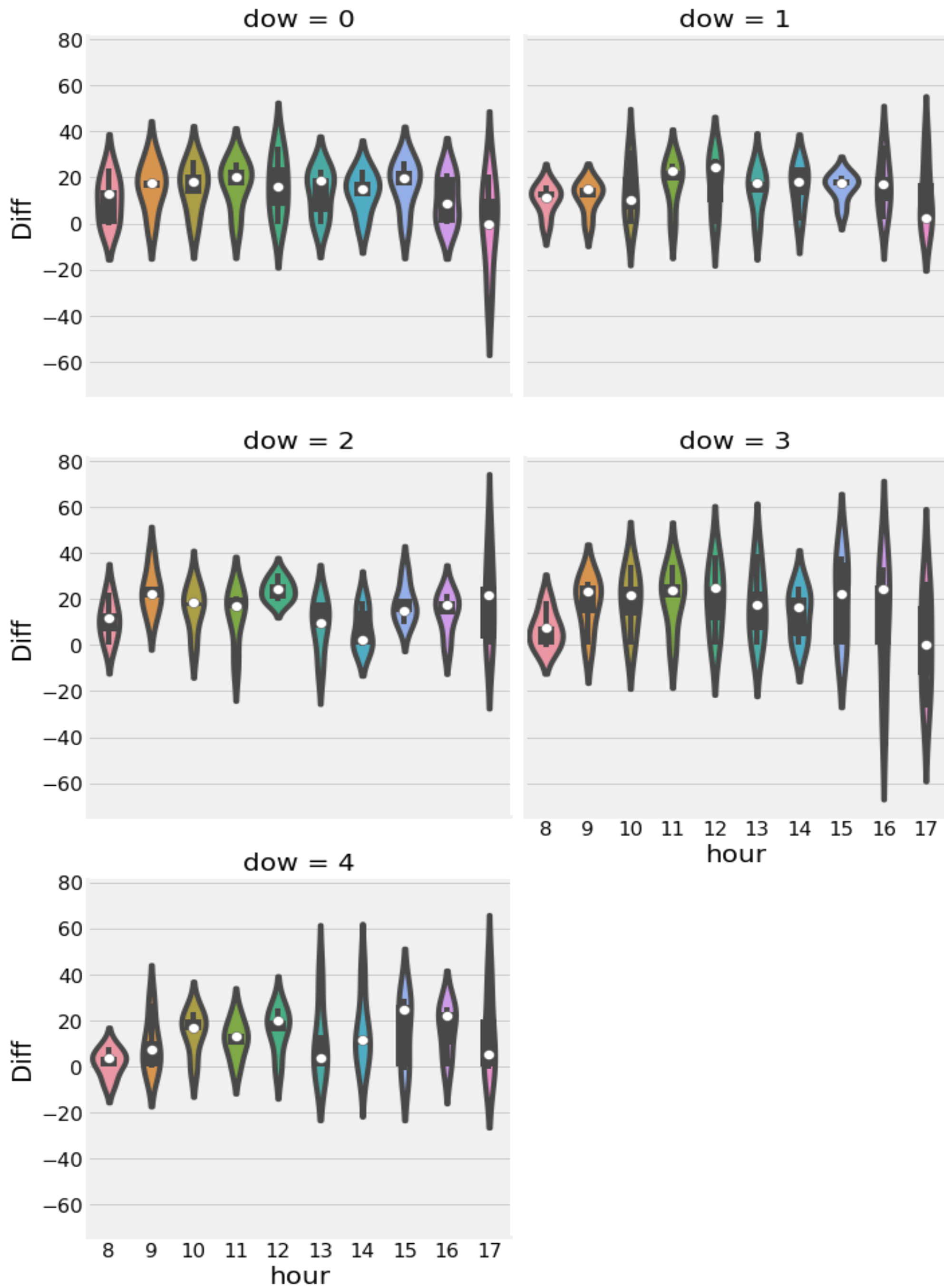

Figure 4.8: Difference of Posture in minutes Vs hours (as per the day of the week)

4.2 Performance Criteria

Consider a univariate time series with feature y. Then y_t is the actual value at time t, $\hat{y}_t$ is the prediction value at time t and m is the number of time steps.

Below are the evaluation metrics equations (4.1 to 4.4) for MAE, MAPE, MSE and RMSE (Vandeput, 2018).

4.2.1 The Mean Absolute Error (MAE)

$$MAE = \frac{1}{m}\sum_{t=1}^{N} | y_t - \hat{y}_t| \qquad 4.1$$

Properties:

- It measures the average absolute deviation of forecasted values from original ones.
- It is also called the Mean Absolute Deviation (MAD).
- Extreme forecast errors are not penalized by the MAE.
- MAE should be as close as possible to 0 for a perfect forecast.

4.2.2 The Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{m}\sum_{t=1}^{N} |\frac{y_t - \hat{y}_t}{y_t}| \qquad 4.2$$

Properties:

- This metric measures the percentage of average absolute error that occurred when forecasting
- Does not penalize extreme forecast errors
- Provide a measure that is independent of the values observed in the series

- Ineffective in data sets where the original measures contain zeros ($MAPE \to \infty$)
- Easily interpretable by people with no statistical background

4.2.3 The Mean Squared Error (MSE)

$$MSE = \frac{1}{m}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2 \qquad 4.3$$

Properties:

- It is a measure of the average squared deviation of the forecasted values
- It penalizes extreme errors during the forecast
- MSE emphasizes that overall forecast error is much affected by large individual errors (outliers)
- It is not easily interpretable metric as the value depends on the dataset and the values present within the Time Series (noisy data or the presence of outliers can hinder the judging of a model's performance)
- This metric is not comparable across all datasets

4.2.4 The Root Mean Square Error (RMSE)

We use the Root Mean Square Error (RMSE) as the evaluation measure. RMSE can be defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_t - y_t\right)}{m}} \qquad 4.4$$

Lower RMSE value implies better accuracy. We use RMSE as the evaluation metric in our study for the following reasons:

- RMSE punishes large errors more than a mean absolute error (MAE) and mean absolute percentage error (MAPE),.
- RMSE avoids using absolute value, unlike MAE. It is highly undesirable, from a computation perspective, to use absolute value in measuring model error sensitivities or for data assimilation applications (Chai and Draxler, 2014)
- RMSE is derived from the sample average, while MAE is derived from the sample median. If data is skewed, then the median falls below the average which is not acceptable. Besides, MAE results in poor forecasts for intermittent data (data with many periods of no demand) (Vandeput, 2018)

4.3 Time-Series Fit Results

Stationary test: Used adfuller test to check the difference posture time series is stationary or non-stationary

Difference Posture time series:

Null hypothesis: Difference posture time series is non-stationary.

Alternative Hypothesis: Difference Posture time series is stationary

As per adfuller test, we get a p-value of 1.85e-16 < 0.05 (95% confidence interval), the null hypothesis has been rejected. Hence data is stationary and has no unit root.

1. Holt-Winter's Method:

Difference Posture:

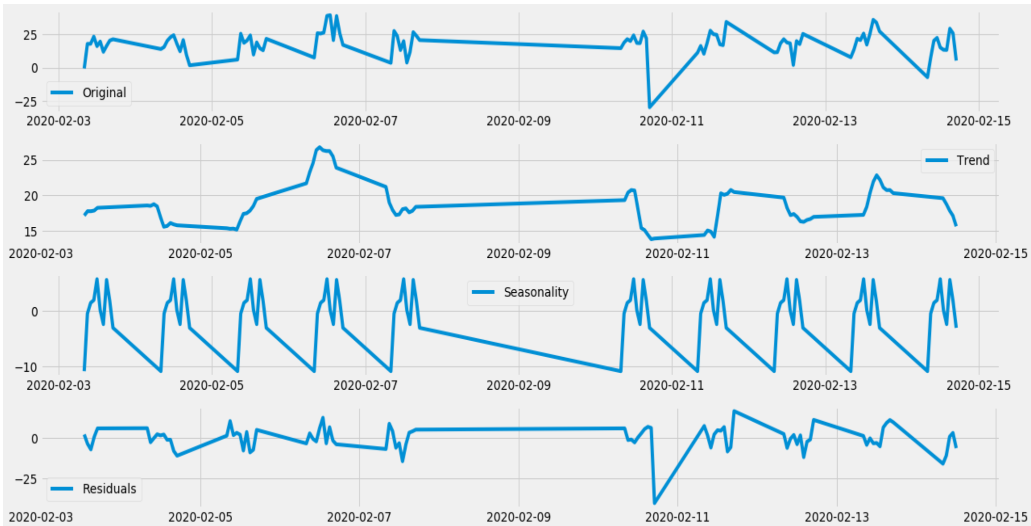

Figure 4.9: Time series components of difference posture in minutes

Using decompose function, we got all-time series components i.e. trend, seasonality and residual for difference posture as shown in Figure 4.9. As we can see that there is no fixed pattern in trend as it keeps on fluctuating in the range from 15 minutes to 30 minutes due to varying time spent by average user for each hour on each day of the week.

As difference posture time series is stationary we can apply Holt winter method and evaluate the performance of the model based on RMSE value.

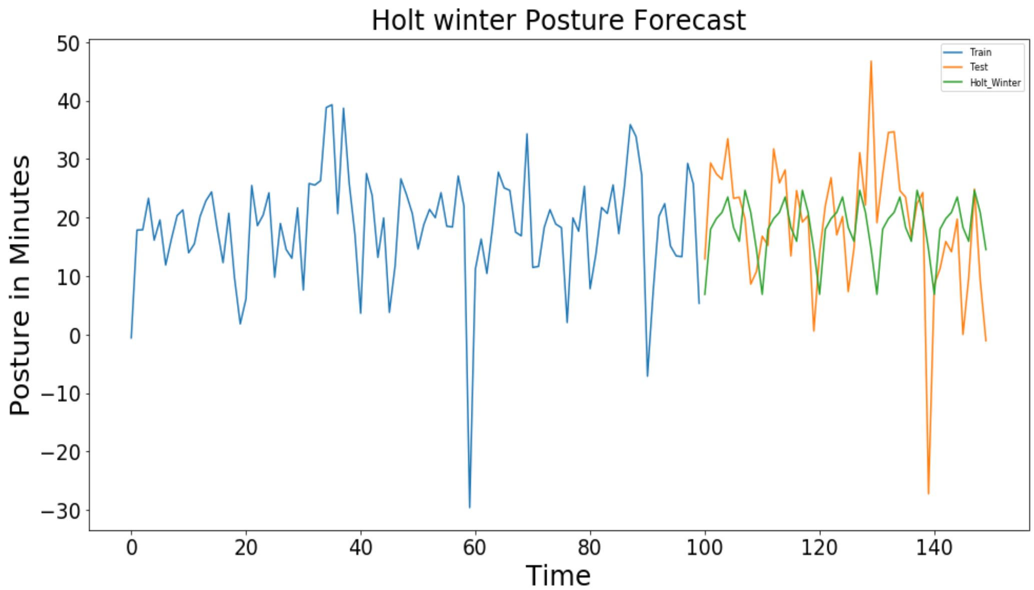

Figure 4.10: Holt-Winter forecast for difference posture time series when each day of the week is same

The Holt-Winter method is trained using two-week data(i.e second and third week) and validated using one week data (i.e fourth week). Here we have assumed that the time spent by the average user is the same for each day of the week and this trained model we named as a global model. The validated result of Holt-Winter method is plotted in Figure 4.10.

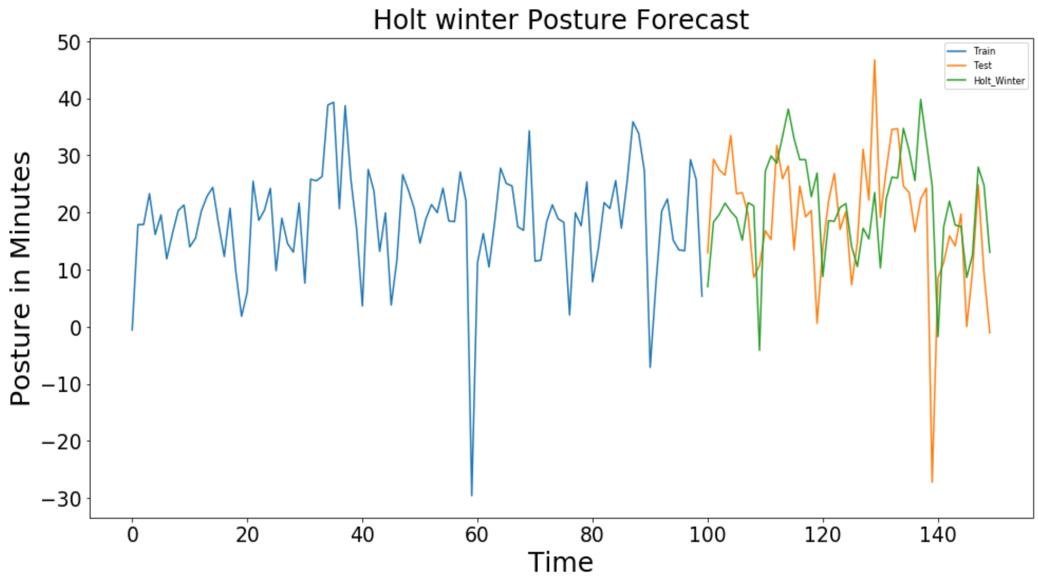

Figure 4.11: Holt-Winter of posture forecast trained for each day of the week.

Here the Holt-Winter method is trained for each day of the week and validated results is merged and plotted in Figure 4.11. The results are evaluated using root mean square error (RMSE) from equation (1) and tabulated in Table 1.1

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Global
08-09	7.03	27.22	8.78	10.29	-1.74	**6.88**
09-10	18.32	29.89	18.56	22.32	17.5	**18.01**
10-11	19.65	28.64	18.47	26.19	21.99	**19.85**
11-12	21.64	33.34	20.9	26.04	17.78	**20.89**
12-13	20.2	38.12	21.56	34.74	17.54	**23.52**
13-14	19.06	32.95	14.03	30.8	8.61	**18.3**
14-15	15.14	29.23	10.52	25.58	12.54	**15.94**
15-16	21.72	29.24	17.24	39.81	27.94	**24.66**
16-17	21.15	22.76	15.36	32.33	24.79	**20.66**
17-18	-4.14	26.91	23.5	24.65	13.01	**14.52**
RMSE	9.42	12.99	9.82	18.82	8.52	**10.85**

Table 4.1: Holt-Winter Method predicted value and RMSE

As we have trained model for each day of the week, the predicted posture value for each hour is different for each day of the week and this results can be compared with the predicted value of Global model (i.e.Global Column). This comparison will help us to guide the users to the rehabilitation program to avoid health-related issues. The user which might get health-related issues are divided into three time shown in Table 4.1 and they are as follows

1. Low-risk region (represented by Orange colour)
2. Medium risk region (represented by Yellow colour)
3. High-risk region (represented by Red colour)

Here we have assumed, Low-risk region= (80% of Global)< posture value < Global value

Medium risk region = (50% of Global value) < posture value < (80% of Global value)

High risk region = posture value < (50% of Global value)

2. ARIMA:

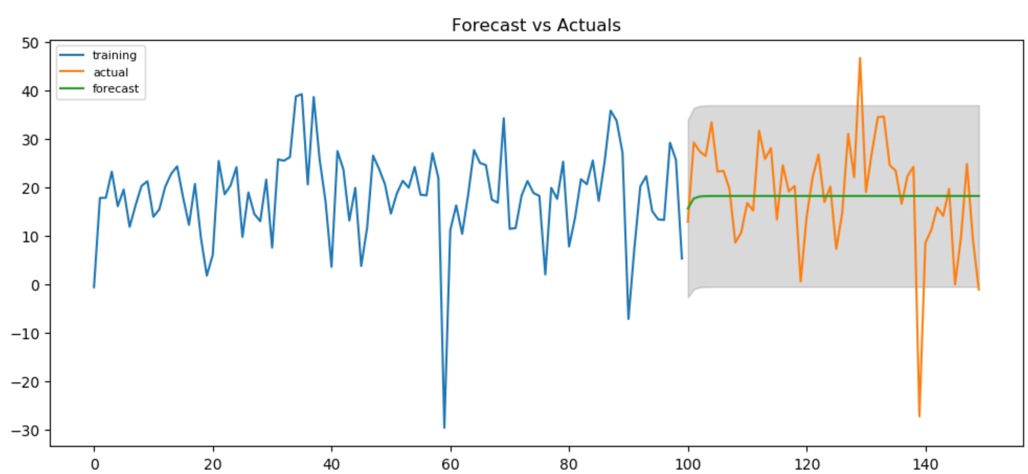

Figure 4.12: ARIMA (p=1,d=0,q=0) forecast for difference posture time series

Here we are tuning the parameter of autoregressive (AR) i.e p, differencing i.e d and moving average i.e q to get the lowest AIC value which will help us to minimize the error. After fitting the ARIMA model consider each day will behave the same as each day of the week i.e Global model. The predicted posture value for test data is plotted in Figure 4.12.

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Global
08-09	-31.11	35.39	14.27	28.25	17.23	**18.26**
09-10	-32.64	36.46	15.64	29.28	17.3	**18.26**
10-11	-34.17	37.53	16.22	30.31	17.36	**18.26**
11-12	-35.7	38.6	16.47	31.35	17.42	**18.26**
12-13	-37.22	39.67	16.57	32.38	17.48	**18.26**
13-14	-38.75	40.74	16.61	33.41	17.54	**18.26**
14-15	-40.28	41.81	16.63	34.44	17.6	**18.26**
15-16	-41.81	42.88	16.64	35.47	17.67	**18.26**
16-17	-43.33	43.95	16.64	36.51	17.73	**18.26**
17-18	-44.86	45.02	16.64	37.54	17.79	**18.26**
RMSE	59.96	22.91	11.84	22.6	9.87	**11.47**

Table 4.2: ARIMA Method predicted value and RMSE

Similarly, here also we have trained the ARIMA model for each day of the week and predicted results are tabulated in Table 4.2. Monday and Tuesday data shows non-stationarity and so there were not enough data points to train the model after converting to stationary data which result in large RMSE value.

Here also we assumed the same assumption for three regions of health-related issues.

3. SARIMA Model:

Similarly, here we tuning the parameters to get the lowest AIC value and predicted the posture for test data and plotted the result shown in Figure 4.13Here we have trained the model assuming each day as same as the day of the week i.e Global model.

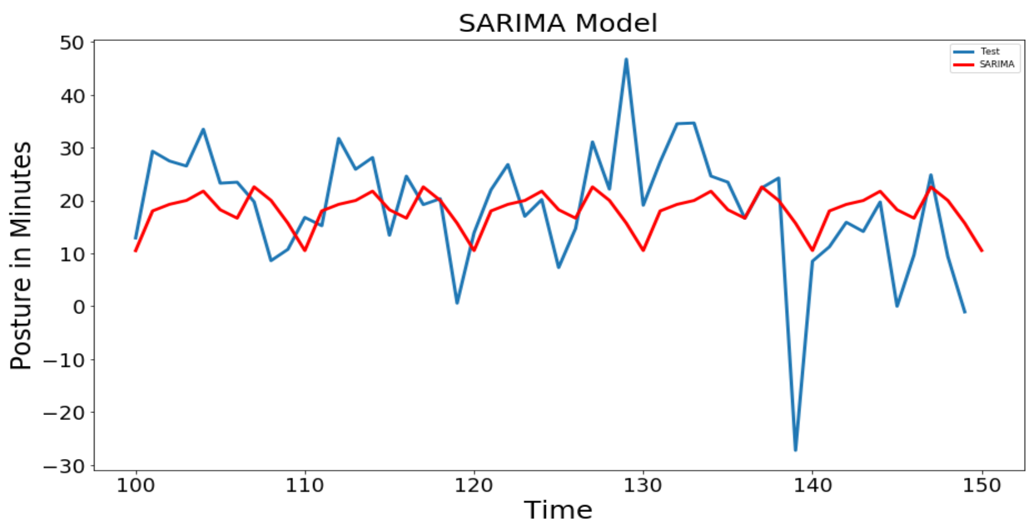

Figure 4.13: SARIMAX(0, 1, 1)x(0, 1, 1, 10) test data Vs Predicted Value of posture

Besides, we trained the SARIMA model for each day of the week and nearly in all cases the model parameter is p=0,d=1,q=1, P=0, D=1, Q=0 with the seasonal period as same as the number of business hour i.e 10. As Monday and Tuesday data was non-stationary we got high RMSE value for test data shown in Table 4.3.

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Global
08-09	-46.64	51.53	10.67	14.78	-15.84	**10.52**
09-10	-42.48	56.63	10.85	20.66	-1.24	**18.04**
10-11	-39.87	50.74	17.52	28.66	11.51	**19.29**
11-12	-41.29	58.76	20.56	27.64	13.65	**20.02**
12-13	-37.02	68.06	18.1	32.53	6.4	**21.78**
13-14	-42.74	65.38	17.46	24.19	4.72	**18.26**
14-15	-42.88	64.94	1.25	32.4	4.6	**16.67**
15-16	-34.16	57.8	19.14	42.83	20.54	**22.59**
16-17	-39.31	57.17	16.87	40.79	17	**20.04**
17-18	-90.86	74.62	24.55	34.16	-3.36	**15.71**
RMSE	68.64	43.12	10.89	22.18	10.38	**10.83**

Table 4.3: SARIMA Method predicted value and RMSE

4. Prophet Method:

Correspondingly, we tune the changepoint_piror_scale parameter of prophet model in range from 0.01 to 0.15 and obtained the Global model results for test data shown in below Figure 4.14

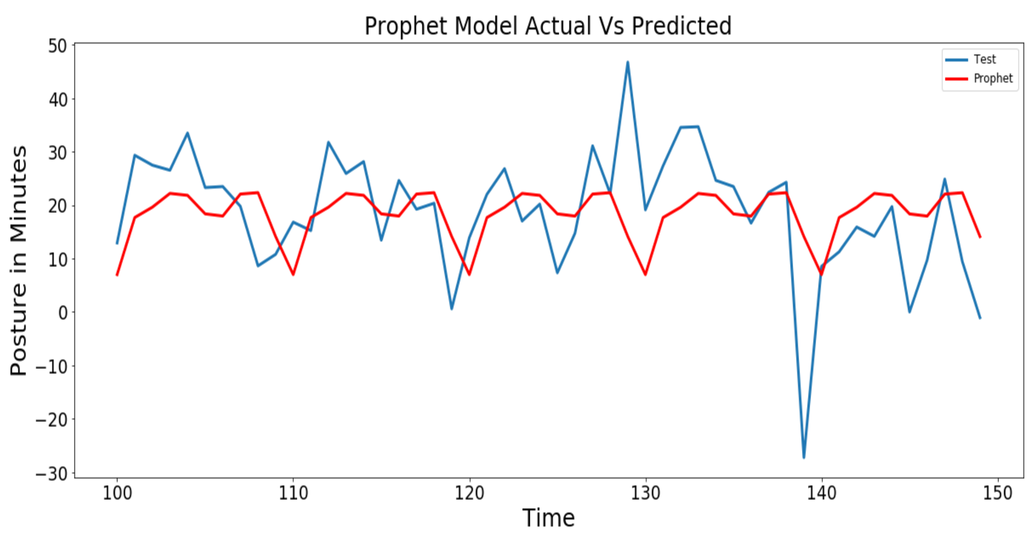

Figure 4.14: Prophet Method Model test data Vs Predicted Value of posture

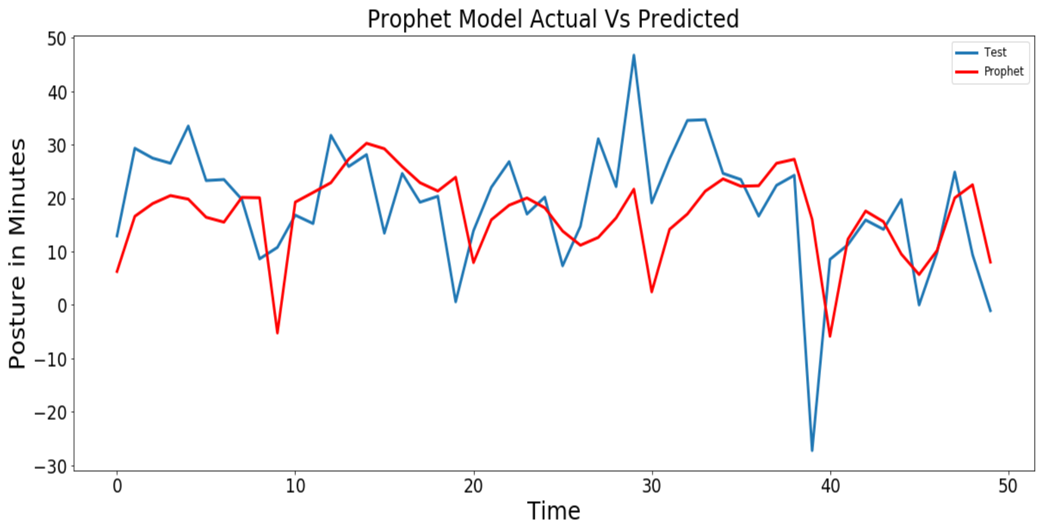

Figure 4.15: Prophet Method of posture forecast trained for each day of the week.

Figure 4.15 shows the results of predicted results for test data by training the prophet model for each day of the week. The predicted results are compared with the global model, the predicted value within the 80 per cent of global value is termed as low-risk time, between 50 per cent to 80 per cent is termed as medium risk time and below 50 per cent of global value is termed as the high-risk time. These regions are marked by colour gradient as shown in Table 4.4.

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Global
08-09	6.24	19.25	7.94	2.46	-5.86	**7.05**
09-10	16.6	21.05	15.94	14.15	12.29	**17.72**
10-11	18.98	22.89	18.68	17.03	17.58	**19.66**
11-12	20.48	27.3	20.02	21.29	15.6	**22.24**
12-13	19.81	30.26	18.19	23.6	9.54	**21.88**
13-14	16.42	29.24	13.84	22.23	5.7	**18.41**
14-15	15.51	25.91	11.19	22.32	10.16	**18**
15-16	20.13	22.89	12.65	26.52	20.04	**22.12**
16-17	20.07	21.33	16.29	27.27	22.5	**22.38**
17-18	-5.24	23.91	21.68	16.01	8.04	**14.18**
RMSE	10	9.66	11	16.93	7.91	**10.9**

Table 4.4 Prophet Method predicted value and RMSE

Comparison of Time series method for posture :

Here we are comparing the MSE, RMSE, MAE.MAPE value of predicted results for test data when we trained the model assuming posture is equal for each day of the week i.e Global model.

Model	MSE	RMSE	MAE	MAPE
Holt-Winters Method	**117.8**	**10.85**	**7.96**	**3248.89**
ARIMA (p=1,d=0,q=0)	**131.57**	**11.47**	**8.41**	**3262.21**
SARIMAX (1, 0, 1)x(1, 0, 1, 10) Model	**117.26**	**10.83**	**7.77**	**3244.56**
Prophet Method	**118.89**	**10.9**	**7.96**	**3255.28**

Table 4.5 Time series model RMSE value (postures same for all weekday)

Based on the lowest value of evaluation metrics, SARIMA model is the best model for posture forecasting when we assumed each day of the week is the same.

Below Table 4.6 shows the RMSE value of Time series method for test data when we trained the model assuming posture is different for each day of the week. As seen from Table 4.2 and 4.3, Monday and Tuesday have very RMSE value so it leads to high RMSE value in ARIMA and SARIMA model. So here we are only comparing Holt-winter and Prophet method.

Model	MSE	RMSE	MAE	MAPE
Holt-Winters Method	**156.24**	**12.49**	**9.544**	**1637.01**
Prophet Method	**132.8**	**11.52**	**8.42**	**1110.27**

Table 4.6: Time series model RMSE value (postures different for all weekday)

Based on the lowest evaluation metrics value, Prophet model is the best model for posture forecasting when we assumed each day of the week is different.

5 Conclusion, Recommendations and Future Scope

Based on the result (Table 4.5)obtained for posture when each weekday is same the error is less (i.e RMSE =10.9 for prophet forecast) compared to the result (Table 4.6) when each day of the week is different (i.e RMSE=11.52 for prophet forecast).So time series method works well when each weekday is assumed as same. So SARIMAX (1, 0, 1)x(1, 0, 1, 10) Model is best model for prediction of posture for the given data.

As per industry partner, each day of the week should be treated as different. So the Prophet method is best for forecasting the posture value for each weekday. Based on Table 4.4, the Company person will be able to guide the employees to the rehabilitation program using the following insights.

On Monday, from 8 am to 1 pm and 4 pm to 5 pm is the low-risk time where the employees are showing a slight decrease in good posture. While from 2 pm to 3 pm is medium risk time where the employees show a significant decrease in productivity. From 5 pm to 6 pm is the high-risk time and during this hour employee's productivity is minimum throughout the day

On Tuesday, nearly on an average, all employees are in a good posture in all hours except from 4 pm to 5 pm where there is a slight decrease in posture. As compared to the other day of the week the employees show the best concentration in work.

On Wednesday, Low-risk time is from 9 am to 12 pm where there is a slight decrease in posture, Medium risk time is from 12 pm to 3 pm and 4 pm to 5 pm where there are moderate decreases in posture. The high-risk time is from 3 pm to 4 pm during this time concentration of employees is minimum.

Similarly, on Thursday, from 8 am to 10 am is medium risk time while from 10 am to 12 am is the low-risk time.

Friday shows employees are least active in work as time spent by employees in bad posture is highest. Nearly all the time employees have low posture value except from 4 pm to 5 pm where employees show good concentration in work. Here High-risk time is from 8 am to 9 am, 12 pm to 1 pm and 5 pm to 6 pm.

In the above insights, we have assumed that 1 pm to 2 pm is lunch hour.

The future work of this study will be extending the number of algorithms used for posture forecasting. such as Recurrent Neural Network (RNN).The time series model will fit improve and minimize the error in future when trained for large posture dataset.

Bibliography

Aronsson, G., Gustafsson, K. and Dallner, M. (2000) 'Sick but yet at work. An empirical study of sickness presenteeism', *Journal of Epidemiology and Community Health*, 54(7), pp. 502–509. doi: 10.1136/jech.54.7.502.

'Back Pain Statistics' (2019) *The Good Body*, 12 February. Available at: https://www.thegoodbody.com/back-pain-statistics/ (Accessed: 11 August 2020).

Box, G. E. P. (2015) 'Time Series Analysis', p. 709.

Bramhapurikar, K. *et al.* (2018) 'A Wearable Posture Corrector Device', in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–5. doi: 10.1109/ICCCNT.2018.8493960.

Brownlee, J. (2016) 'How to Check if Time Series Data is Stationary with Python', *Machine Learning Mastery*, 29 December. Available at: https://machinelearningmastery.com/time-series-data-stationary-python/ (Accessed: 22 August 2020).

Carpenter, B. *et al.* (2017) 'Stan: A Probabilistic Programming Language', *Journal of Statistical Software*, 76(1), pp. 1–32. doi: 10.18637/jss.v076.i01.

Chai, T. and Draxler, R. R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature', *Geoscientific Model Development*, 7(3), pp. 1247–1250. doi: 10.5194/gmd-7-1247-2014.

Dr. Iain Pardoe (2018) *14.1 - Autoregressive Models | STAT 501*, *PennState: Statistics Online Courses*. Available at: https://online.stat.psu.edu/stat501/lesson/14/14.1 (Accessed: 22 August 2020).

Epstein, R. *et al.* (2012) 'The effects of feedback on computer workstation posture habits', *Work*. IOS Press, 41(1), pp. 73–79. doi: 10.3233/WOR-2012-1287.

Gardner, E. S. (1985) ‘Exponential smoothing: The state of the art’, *Journal of Forecasting*, 4(1), pp. 1–28. doi: 10.1002/for.3980040103.

Ghanbari, A. *et al.* (2008) ‘Effect of forward shoulder posture on pulmonary capacities of women’, *British Journal of Sports Medicine*. British Association of Sport and Excercise Medicine, 42(7), pp. 622–623. doi: 10.1136/bjsm.2007.040915.

Goossens, R. h. m., Netten, M. p. and Van der Doelen, B. (2012) ‘An office chair to influence the sitting behavior of office workers’, *Work*. IOS Press, 41, pp. 2086–2088. doi: 10.3233/wor-2012-0435-2086.

van den Heuvel, S. G. *et al.* (2007) ‘Loss of Productivity Due to Neck/Shoulder Symptoms and Hand/Arm Symptoms: Results from the PROMO-Study’, *Journal of Occupational Rehabilitation*, 17(3), pp. 370–382. doi: 10.1007/s10926-007-9095-y.

Kang, K. W. *et al.* (2016) ‘Effect of sitting posture on respiratory function while using a smartphone’, *Journal of Physical Therapy Science*, 28(5), pp. 1496–1498. doi: 10.1589/jpts.28.1496.

Lin, F. *et al.* (2006) ‘Effect of Different Sitting Postures on Lung Capacity, Expiratory Flow, and Lumbar Lordosis’, *Archives of Physical Medicine and Rehabilitation*. Elsevier, 87(4), pp. 504–509. doi: 10.1016/j.apmr.2005.11.031.

mainpath (2017) *Why Everyone Keeps Saying ‘Sitting Is the New Smoking’*, *Tri-City Medical Center*. Available at: https://www.tricitymed.org/2017/07/everyone-keeps-saying-sitting-new-smoking/ (Accessed: 21 August 2020).

Nair, S. *et al.* (2015) ‘Do slumped and upright postures affect stress responses? A randomized trial’, *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, 34(6), pp. 632–641. doi: 10.1037/hea0000146.

Nau, R. (2014) ‘Mathematical structure of ARIMA models’, p. 8.

Nobari, M. *et al.* (2017) ‘Effect of Corrective Exercises on Cervicogenic Headache in Office Workers With Forward Head Posture’, *Journal of Modern Rehabilitation*, 11(4), pp. 201–208.

Rob J Hyndman (2010) *Initializing the Holt-Winters method |*. Available at: https://robjhyndman.com/hyndsight/hw-initialization/ (Accessed: 13 August 2020).

Rob J Hyndman (2013a) *6.1 Time series components | Forecasting: Principles and Practice*. Available at: https://Otexts.com/fpp2/ (Accessed: 23 August 2020).

Rob J Hyndman (2013b) *8.1 Stationarity and differencing | Forecasting: Principles and Practice*. Available at: https://Otexts.com/fpp2/ (Accessed: 22 August 2020).

Rob J Hyndman (2013c) *8.4 Moving average models | Forecasting: Principles and Practice*. Available at: https://Otexts.com/fpp2/ (Accessed: 22 August 2020).

van Rossum, G. (1995) 'Python tutorial'. Available at: https://ir.cwi.nl/pub/5007 (Accessed: 14 August 2020).

Russell, H. *et al.* (2018) '485 The trends and determinants of work-related musculoskeletal disorders (msd) in ireland between 2002 and 2013', *Occupational and Environmental Medicine*. BMJ Publishing Group Ltd, 75(Suppl 2), pp. A270–A271. doi: 10.1136/oemed-2018-ICOHabstracts.773.

Schrempf, A. *et al.* (2011) 'PostureCare – Towards a novel system for posture monitoring and guidance', *IFAC Proceedings Volumes*, 44(1), pp. 593–598. doi: 10.3182/20110828-6-IT-1002.02987.

Stephanie (2019) *ARMA model*, *Statistics How To*. Available at: https://www.statisticshowto.com/arma-model/ (Accessed: 22 August 2020).

Taylor, S. J. and Letham, B. (2018) 'Forecasting at Scale', *The American Statistician*. Taylor & Francis, 72(1), pp. 37–45. doi: 10.1080/00031305.2017.1380080.

Van Eerd, D. *et al.* (2016) 'Effectiveness of workplace interventions in the prevention of upper extremity musculoskeletal disorders and symptoms: an update of the evidence', *Occupational and Environmental Medicine*, 73(1), pp. 62–70. doi: 10.1136/oemed-2015-102992.

Vandeput, N. (2018) ‘Forecast KPI: Bias, MAE, MAPE & RMSE’, *SupChains*. Available at: https://supchains.com/article/forecast-kpi-bias-mae-mape-rmse/ (Accessed: 23 August 2020).

Wang, W.-J. *et al.* (2016) ‘Human Posture Recognition Based on Images Captured by the Kinect Sensor’:, *International Journal of Advanced Robotic Systems*. SAGE PublicationsSage UK: London, England. doi: 10.5772/62163.

Womersley, L. and May, S. (2006) ‘Sitting posture of subjects with postural backache’, *Journal of Manipulative and Physiological Therapeutics*, 29(3), pp. 213–218. doi: 10.1016/j.jmpt.2006.01.002.

Xu, W. *et al.* (2013) ‘eCushion: A Textile Pressure Sensor Array Design and Calibration for Sitting Posture Analysis’, *IEEE Sensors Journal*. *IEEE Sensors Journal*, 13(10), pp. 3926–3934. doi: 10.1109/JSEN.2013.2259589.

Zhang, Z. *et al.* (2014) ‘A novel method for user-defined human posture recognition using Kinect’, in *2014 7th International Congress on Image and Signal Processing*. *2014 7th International Congress on Image and Signal Processing*, pp. 736–740. doi: 10.1109/CISP.2014.7003875.