**CHAPTER THREE**

**METHODOLOGY**

**3.1 Data Presentation and Collection Procedure**

The data is a secondary data that was retrieved from Yahoo Finance API using python. First, you install the package ***yfinance***  using the command

**pip install yfinance**

Afterwards you run the following code to retrieve the data

**import yfinance as yf**

**data = yf.download(“AAPL”, start\_date= “2015-01-01”, end\_date=“2023-12-01”, interval= “1mo”)**

The data is a apple stock that consists of variables such as date which has a time interval of one month, open which is the open price index of the stock market, close which is the close price index, Adj close which is the adjusted close price index and volume which is the market volume.

**3.2 Prophet Model**

The prophet model is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

The prophet model uses a decomposed times series model with three main components: trends, seasonality, holidays. They can be combined into an equation written as:

*Y(t) = g(t) + s(t) + h(t) + et*

**(I)**

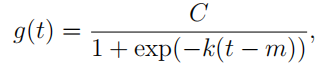
Where *g(t)* is the trend function that models non-periodic changes, *s(t)* is the periodic changes (e.g weekly, monthly or yearly seasonality), *h(t)* is the effect of holiday which records one or more irregular schedules and *et* is the idiosyncratic changes which are not accommodated by the model.

This specification is similar to a generalized linear model (GAM) (Hastie & Tibshirani, 1987), a class of regression models with potentially non-linear smoothers applied to the regressors. In this scenario, time is used as the regressor but possibly several linear and non-linear function of time as components. Modeling seasonality as an additive component is the same approach taken by exponential smoothing (Gardner, 1985).

*3.2.1 Trend Model*

There are two trend models that can be implemented in the prophet model: saturating growth model and piecewise linear model. The trend model that will be used for analysis depends on the structure of the time series data.

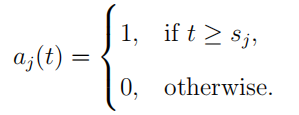
For growth forecasting, the core component of the data generating process is a model for how the population has grown and how it is expected to continue growing. Modeling growth is often sometimes similar to population growth in natural ecosystems, where there is nonlinear growth that saturates at a carrying capacity. For example, the carrying capacity for the number of Facebook users in a particular area might be the number of people that have access to the Internet. This sort of growth is typically modeled using the logistic growth model, which in its most basic form is



**(II)**

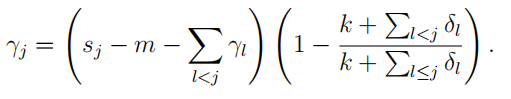
with *C* the carrying capacity, *k* the growth rate, and *m* an offset parameter.

We incorporate trend changes in the growth model by explicitly defining changepoints where the growth rate is allowed to change. Suppose there are *S* changepoints at times *sj* , *j* = 1*, . . . , S*. We define a vector of rate adjustments ***δ*** *∈* R*S* , where *δj* is the change in rate that occurs at time *sj* . The rate at any time *t* is then the base rate *k*, plus all of the adjustments up to that point: *k* + **Σ***j*:*t>sj  δj .* This is represented more cleanly by defining a vector **a**(*t*) *∈ {*0*,* 1*} S* such that



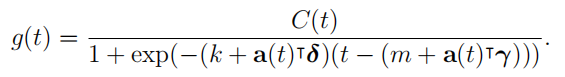
**(III)**

The rate at time *t* is then *k* + **a**(*t*) | ***δ***. When the rate *k* is adjusted, the offset parameter *m* must also be adjusted to connect the endpoints of the segments. The correct adjustment at changepoint *j* is easily computed as



**(IV)**

The piecewise logistic growth model is

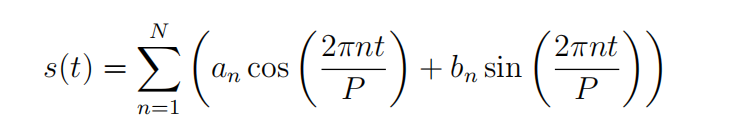


**(V)**

An important set of parameters in our model is *C*(*t*), or the expected capacities of the system at any point in time. The logistic growth model presented here is a special case of generalized logistic growth curves, which is only a single type of sigmoid curve. Extensions of this trend model to other families of curves is straightforward.

*3.2.2 Seasonality*

Business time series often have multi-period seasonality as a result of the human behaviors they represent. For instance, a 5-day work week can produce effects on a time series that repeat each week, while vacation schedules and school breaks can produce effects that repeat each year. To fit and forecast these effects we must specify seasonality models that are periodic functions of *t*. We rely on Fourier series to provide a flexible model of periodic effects (Harvey & Shephard 1993). Let *P* be the regular period we expect the time series to have (e.g. *P* = 365*.*25 for yearly data or *P* = 7 for weekly data, when we scale our time variable in days). We can approximate arbitrary smooth seasonal effects with



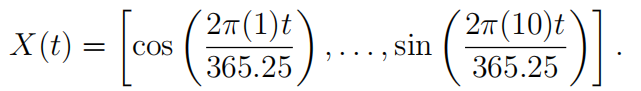
**(VI)**

a standard Fourier series. Fitting seasonality requires the *2N* parameters. This is done by constructing a matrix of seasonality vectors for each value of *t* in the times series historical and future data.

**B** = [*a1,b1, …, aN,bN*]T

**(VII)**

For example, a data with yearly seasonality and  *N = 10* can be written as:



The seasonal component can be written as:

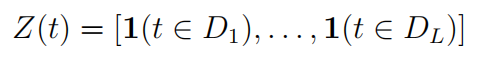
*s(t) = X(t)****B***

**(VIII)**

*3.2.3 Holidays*

Holidays and events provide large, somewhat predictable shocks to many business time series and often do not follow a periodic pattern, so their effects are not well modeled by a smooth cycle.

Incorporating this list of holidays into the model is made straightforward by assuming that the effects of holidays are independent. For each holiday *i*, let *Di* be the set of past and future dates for that holiday. We add an indicator function representing whether time *t* is during holiday *i*, and assign each holiday a parameter *κi* which is the corresponding change in the forecast. This is done in a similar way as seasonality by generating a matrix of regressors



**(IX)**

and taking

*h(t) = Z(t)****k***

**(X)**

It is often important to include effects for a window of days around a particular holiday, such as the weekend of Thanksgiving. To account for that we include additional parameters for the days surrounding the holiday, essentially treating each of the days in the window around the holiday as a holiday itself.

**3.3 Performance Metrics**

The accuracy of the model will tell how accurate the forecast will be and error in the forecast. This can be achieved through the mean square error(MSE), root mean square error(RMSE), root mean square percentage error(RMSPE) of the model.

The MSE is the average difference between the predicted and actual value in a model and is given by

The RMSE is the magnitude error of the predicted and actual value. The RMSE is given by

The RMSPE is percentage error between the predicted and actual value and is defined by