Present Learning of the Newspaper Circulation and Sales Forecasting

2.1 Introduction

Previous chapter provide indirection for this research project. It's outlined the project aims and objectives as well as brief description about propose solution. This chapter is focused to review previous work done in this research domain.

2.2 Background Literatures

A newspaper is mass media relatively inexpensive daily publications which include important news, articles, business information, advertising, entitlement and other information. Newspapers classify based on the publication as daily or weekly and circulation strategies based on this classification. Newspapers can be classified as daily, weekly, national, international, and online and customize and various interest newspapers (Young, Woman, Children, Sport). National newspapers play significant role in the particular country because it is serving the country and region as well as these newspapers have large customer base compare to other newspapers. Large number of employees' work daily or weekly, in national newspaper establishments, and it need to be up to date frequently. Therefore, national newspapers incurred high cost. Newspaper publication business, is owns by stakeholders. Stakeholders of the newspaper industry are expecting profitability. Newspaper circulation and advertising is main income source of the newspaper industry [1]. Achieving high profitability, industry need to, running smooth process for circulate their publication effectively. Companies need to distribute correct amount of newspapers to their customers and deliver their sales point at right time without any shortage.

In written history great Emperor Julies Caesar published the government bulletins before the 17th century [1]. First Newspaper "*Avisa Relation olde Zeitung*" published in Germany in the early 17th century. In 1662 first English language newspaper "*The Weekly news*" published in England [4]. Evolutions of the printing technology

newspapers become cheaper and wider available method to communicate the information to the society. Due to technology development newspaper industry popularity was simultaneously improved [1]. Archived the demand of newspapers, publication companies improve their newspaper circulation process efficiency. Therefore, they introduce circulation system to the industry.

Sri Lankan newspaper industry was initiated in colonial era as a recommendation of Colebrook Commission in 11th April 1829. In 1832 *"Colombo Journal"* first Sri Lankan newspaper was published during Governor Wilmot Horton's period. The first Sinhala newspaper, *"Lanka Loka"* was published in Galle in June 1860. The first Tamil Newspaper *"Idea Atari"* was published in 1841 [5].

Today twenty four national newspapers and more than seventeen other newspapers are published in Sri Lanka. National newspapers can classified as ten Sinhala newspapers, five Tamil newspapers and nine English newspapers [3]. According to the Central Bank of Sri Lanka in year 2013 daily national newspaper circulation is nearly 225,528,000 and weekly national newspaper circulation is nearly 103,572,000 [2]. There is a high demand for newspapers in Sri Lankan society. The newspaper publishing companies now adopt the circulation systems and sales forecasting systems as companies target high profitability. Therefore, companies looking for forecasting systems for sales by maximum circulation with less return. .

Business planning and decision making, forecasting plays a significant role. Forecasting is a most important and first stage of the business planning. Accurate forecasting is helps to reduce the risk and ensure the success of the decision making. Newspaper publishing companies forecast the most favorable supply of newspapers that minimize the cost and maximize the profit under the vague newspaper demand with minimum returns. If company supply more than the requirement or demand there will be unnecessary costs such as return costs, stocking, distribution and supply in shortage, company faced sales lost [6].

Incesu at al developed short term forecasting system for predict newspaper net sales by using time series analysis [6]. In this solution they takes records of past time intervals and forecast future time intervals. Incesu at al based to use Gramger and Newbold and Reid finding that including time series methods, exponential smoothing and ARIMA for forecasting [7]. However, Incesu at al findings is difficult to predict

sales increase suddenly situations like special event occurs or special promotional periods (non leaner approach) etc; .

Interim Report "US Newsprint Demand Forecasts to 2020" combining the several forecasting methods and try to predict the newsprint demand in US [8]. First they used Classical model and forecast the structural changes of the newsprint usage patterns by using historical data. Then, they study about Gross Domestic Product (GDP) and newsprint demands. Then, they look to Bayesian Module that accepts to used industry expert knowledge about future demand of the newsprint and predict the estimation. Ad hoc newspaper circulation model used for calculate the predication by using function of circulation that effect to newsprint demand. However, classical model fails to forecast the structural changes in newsprint demand.

'Just Enough Delivery (JED)' forecasting system that develop by Heskes et al helps for optimizing the newspaper sales in the companies [9]. This system used for combination of neural and Bayesian methodology. Bayesian helps to figured out the similarities between different sales points and try to forecast the demand in unpredictable market changes. System is robust against unpredictable situations helps of correction terms in prediction model. However, limitation of the data availability this research default to identified the relationships (non linier or linier) between the parameters in time series.

Calli and Weverbergh using censored regression technique for forecasting the newspaper demand [10]. Large fraction of observations at the minimum or maximum had censored variable. It's a natural approach for newspaper sales furcating and it provide information on key variables and its simple to link with the distribution policy as well as it easy to adopt the existing forecasting and distribution policy. The results of the outcomes are more favorable for comparing traditional moving average approach. It's more suitable for non liner sales and returns situation in each single newspaper sales outlets [10],[11],[12]. However, this method allowed a straightforward evaluation and applies for all sales points. It not considers about differentiation of the each sales points as well as sudden change situation of the demand.

Forecast the number of users who visit one website domain, Napagoda used data mining techniques that used for time series data forecasting [13]. Use of result researcher analyzed and forecast time dependent data point. Analyzed the time series, these research used Weka (Waikato Environment for Knowledge Analysis) data mining software version 3.7.6., which support to time series sales forecasting environment. This research used two type of data set, one data set contain outlier data and other data set remove outliers. Outlier means sudden changers of data occasionally. In this analysis, this research use following classifiers that contains in Weka such as Gaussian Process, Multilayer perceptron, Linear Regression and SMO Regression. After analyzed this research evaluated the forecast results based on Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Mean Square Error (RMSE) and Relative Absolute Error (RAE) which available in Weka tool. Research concludes SMO Regression and Linear Regression algorithms are more suitable to forecast the web site related information. Therefore, Newspaper sales forecasting Weka tool can be used. However, Napagoda's research findings based on short time period of past data set. Therefore, accuracy of forecasting results makes doubtful.

Similar to above research Kumar and Balara use Weka Tool version 3.7.8 times series forecasting module for forecast the Indian stock market NIFTY index in next ten days [14]. This research used past three month stock market values. This research predicts Market open value, close value, high and low value of particular date by using history data. Data preprocessing researchers used minimum, maximum mean and standard deviation in previous mention attributes. Then this research used stock values lies in between standard deviation range. This research study and used Jena and Swasti 's findings in weka tool usage on data preprocessing, classification and clustering [15]. Then researchers applied Gaussian process, Linear regression, Multilayer perceptron and SMO regression methods on three month past data forecast the values. These methods are already available in Weka tool. Then researchers are compared results with actual and error percentage. Consider evaluation researchers conclude SMO regression is more suitable method to forecast the stock price. Sureshkumar and Elango also did the similar kind of research for forecast the Indian stock market by using Weka tool [16]. However, Kumar and Balara used only three month data,

increase the accuracy of forecasting results researchers must used more historical data.

Kannan et al forecast the rainfall by using data mining techniques [17]. They used data mining techniques such as clustering, classifying, multiple linear regression method used for forecast the rainfall in the future and measures the rainfall they used Pearson correlation coefficient. Kannan et al based on findings in Guhathkurta that contain on time series data and natural networking model [18] . During the testing they figure out some results vary from actual because of uncertainty of the weather.

Perez at al studied about the natural network forecasting methods such as RBFN and CO2RBFN for short term forecasting of extra virgin olive oil price in the Spanish Market [19]. Fist researchers analyzed olive oil price time series means of ARIMA method. This analysis shows the non stationary of time series and the convenience of the differentiate it. The results of CO2RBFN were compared with ARIM models and the data mining methods such as a fussz system developed with GA-P etc; the results output most equal but slightly different form data mining method and ARIMA method.

This above study shows numerous limitations of forecasting related to newspaper sales as well as similar areas. Among other issues are, selecting the accurate forecasting methods, sudden changes in sales demand, demand changers in client point and client geographical locations can be highlighted. These issues are summarized in table 2.1.

Research	Limitation
Sales Forecasting System for Newspaper Distribution Companies in Turkey- Incesu at al	Difficult to predict sales increase suddenly situations like special event occurs or special promotional periods.
"US Newsprint Demand Forecasts to 2020"- Hetemäki, L., Obersteiner	Findings fail to forecast the structural changes.
Bayesian learning for sales rate prediction for thousands of retailers- Heskes et al	Limitation of the data availability this research default to identified the relationships (non linier or linier) between the parameters in time series.
Forecasting newspaper demand with censored regression -Calli and Weverbergh	• Forecasting method allowed a straightforward evaluation and applies for all sales points. • It not considers about differentiation of the each sales points as well as sudden change situation of the demand.
Web Site Visit Forecasting Using Data Mining Techniques- Napagoda, C	Research findings based on short time period of past data set.
Time Series Forecasting Of Nifty Stock Market Using Weka -Kumar, R., Balara, A.	Research findings based on short time period of past data set.
Forecast the Rainfall by Using Data Mining Techniques- Kannan et al	Some results vary from actual because of uncertainty of the weather.
An Study on data mining method for Short- term forecasting of the extra virgin olive oil price in the Spanish Market- Perez et al.	This finding use for short term Forecasting. .

Figure 2. 1 Summarized Forecasting issues in Literatures

Based on the above, the research problem can be defined as how to design the accurate sales forecasting system for newspaper net sales.

2.3 Background of the Company

This research project based on The Associated Newspapers of Ceylon limited (ANCL) is a one of the largest and pioneer newspaper publication company in Sri Lanka. During the first decade of the 19th century, late Mr. D.R. Wijewardena founded this company [5]. Currently, company publishes eleven newspapers and magazines. Out of the eleven newspapers, there are 6 national newspapers that cover 3 national languages. "Silumina","Dinamina" for Sinhala readers, "Sunday Observer" , "Daily News" for English readers and "Thinakaran", "Thinakaran Varamanjari" for Tamil readers [20].

ANCL has island wide circulation coverage. ANCL has more than 1087 registered news agencies and under the main agencies there are several sub dealers. They are groceries, railway stations, markets etc; Other than agencies company has six branch offices. Also they have twenty one Regional Sales Executives, Street Sellers, Miscellaneous Customers and Subscriptions.

2.4 Summary

Literature available in newspaper circulation and forecasting are less in Sri Lankan context. In sales forecasting most of the researchers used time series analysis and some of them used other techniques. Some researchers use Weka data mining software and forecast the results. In time series forecasting most of research use following classifiers such as Gaussian Process, Multilayer perception, Linear Regression and SMO Regression. These algorithms already build in the Weka data mining software. Some researchers consider only sales of the newspapers. They not consider the return amount of the paper. Consider net sales (different between issue and returns) only get accurate sales forecast in newspapers. Next chapter consider about adopted technologies that used for solve the research problem.