A Study of Sentiment Analysis and Sales Prediction: Tourism Domain

Odilia Gonsalves
Department of Information Technology
Atharva College of Engineering
Mumbai, India

Abstract-In this study, I aim to predict sales for a tourism sector based on the mined reviews. To validate the forecasting model, real world data from a tourism firm is used. The algorithm for sales prediction is validated and compared against other time series models. The performance of Holt’s Winter method is compared with the performance of other time series techniques. Results obtained from sales prediction show that Holt’s Winter Method performs better than the other time series forecasting techniques. The sales prediction is calculated as a combination of sentiment analysis score and Holt’s winter method. The derived result states that the forecasted value is varied based on the sentiment analysis score for past months. The conclusion is that for a tourism sector having seasonal trend, Holt’s Winter Method in combination with the predictive power of reviews performs better than the other time series forecasting techniques. In future, the same data can be used to predict customer behavior and utilize it for improving the company’s business can be implemented. Also different algorithms like ARSA can be applied on the current dataset and their accuracy in forecasting can be calculated.

Keywords:- Sentiment Analysis, Sales Prediction, Time Series Forecasting

I. INTRODUCTION

The most commonly used text classification tool that analyses an incoming message and specifies whether the sentiment expressed in that message is positive or negative is Sentiment Analysis. It is also known as Opinion Mining. Forecasting can be defined as the act of giving advance warning in time for beneficial actions to be taken. The process of predicting future sales is known as Sales Forecasting. Sales forecasting helps to achieve sales goals like driving sales revenue, improving efficiency and increasing customer retention. Reviews are very important in today’s world. They are considered as a form of customer feedback. Internet is used by everyone today. People give their opinion regarding a product on web in the form of reviews. Reviews can influence tourism industry. Only classifying reviews as positive or negative are of no use. The hidden factors in sentiments of the reviews are to be analyzed. Another factor I consider in this research is sales prediction based on past sales performance of the tourism company. Thus, in this paper, my aim is to explore the predictive power of reviews in the tourism domain and predict sales using sentiment information mined from reviews. The paper is organized as follows:

II. LITERATURE SURVEY

A. Nisha Jebaseeli, E. Kirubakaran. [3] implemented simpler version of the sentiment-aware auto-regressive model and found that this model can produce very good performance for predicting the box office sale revenue using online review data. Gautam Tripathi, Naganna S. [4] presented a survey on sentiment analysis and the related techniques. They also discussed the application areas and challenges for sentiment analysis with insight into the past researches. Gurudeo Anand Tularam, Tareq Saeed, [5] described that Exponential Smoothing (ES), Holt-Winters (HW) and autoregressive integrated moving average (ARIMA) models were compared and found that HW model performed better. Prajakta S. Kalekar. [6] analysed seasonal time series data using Holt-Winters exponential smoothing methods. Rui Yao, Jianhua Chen., [7] used sentiment analysis and machine learning methods to study the relationship between the online reviews for a movie and the movie’s box office revenue performance. Samaneh Beheshhti-Kashia et.al. [8] presented methods in the sales forecasting research with a focus on fashion and new product forecasting.
III. PROPOSED SYSTEM

The Proposed Architecture consists of following layers:

1. Presentation Layer - GUI composes of a web application that will provide the facility to request sales prediction for tourism industry. The result will be displayed in tabular/graphical format.

2. Business Logic Layer - This composes of a business logic component. This component will process the reporting data and send it to the presentation layer.

3. Data Access Layer - This consist of a database access components used for retrieving data from database.

3.1 Execution flow for the Proposed System
Flow for the proposed system is as shown in the Figure 3.1

3.2 UML Diagrams

Use Case Diagram

Figure 3. Block diagram for proposed system

Figure 3.1. System flow diagram

Figure 3.2.1 Use Case Diagram
3.3 UI Design

Tourism Company Website Home Page

Figure 3.3.1 Home Page of Trendy Travel

Login Screen for User and Admin

Figure 3.3.2 User and Admin Login Form

Admin Module Description

Login: In order to access the system, admin need to login first using valid id and password.

Add Places: Admin can add planned tours.

Add Reviews: Admin can upload an Excel Sheet, which will consist of Date, Review and Rating. It also provides admin the facility to upload past sales data.

View Reviews: Admin can view the sentiment analysis score generated for the reviews.

View Sales Report: Admin will be shown a Graph which will show sales predicted based on the past sales data and sentiment analysis.

Add Places Screen: (Admin only)

Figure 3.3.3 Add Places Form for Admin

System provides facility to upload Review files. Accessible only to Admin.
System provides facility to Admin to view the ratings provided to reviews, for a particular place by selecting Month and Place ID.

System provides facility to Admin to view Sales Report by selecting a place Id.

User can have a look on different Packages added by admin and view comments on them.

System provides facility to the User to send feedback to Admin.

IV. IMPLEMENTATION ALGORITHMS

Implementation of Sentiment Analysis Algorithm

Step 1: Upload reviews into the Database.
Step 2: Read reviews for the duration defined.
Step 3: For each review, compare the word in reviews against the dictionary.
Step 4: Analyze the positive/negative sentiments and generate a score.
Step 5: Store the score against the reviews.
Step 6: Repeat steps 3 to 5 till the last review is analyzed.

Implementation of Sales Forecasting Algorithm

Step 1: Upload the past sales data into the database.
Step 2: Implement the sales forecasting algorithm on the uploaded data.
Step 3: Take average score of the sentiment analysis done.
Step 4: Apply a probabilistic approach to influence the predicted sales based on sentiment analysis.
V. RESULTS AND DISCUSSIONS

Experimentation

The sample data containing reviews for tourism sector is taken for experimentation.

<table>
<thead>
<tr>
<th>Date</th>
<th>TourID</th>
<th>Tour</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>8/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>8/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>8/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>8/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>9/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>9/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>9/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>9/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
<tr>
<td>9/15/2017</td>
<td>101</td>
<td>Singapore Holiday</td>
</tr>
</tbody>
</table>

Figure 5.1 Review data for Tourism industry

The data is being processed and Sentiment Analysis is applied on this data to derive a positive/negative score as shown in Figure 5.2

Figure 5.2 Sentiment Analysis score

This data is provided to the forecasting algorithm to predict sales as shown in figure 5.4

Figure 5.3 Sales data for Tourism industry

Figure 5.4 Sales Prediction for Tourism industry

Figure 5.4 shows the performance of Holts Winter forecasting method along with Mean Absolute Deviation, Mean Squared Error and Mean Absolute Percentage Error.
Figure 5.5 shows the graphical representation of the Figure 5.4 data. It also shows the prediction varied due to the results of sentiment analysis.

Comparison of Time Series Forecasting Techniques

This section shows comparison between performances of various Time Series forecasting techniques.

Forecasting using Simple Moving Average

Forecasting using Weighted Moving Average

Forecasting using Exponential Smoothing

Forecasting using Adaptive Rate Smoothing/Holt’s Winters Forecasting

Discussions

From the above derived results, the analysis shows that Holt’s Winter Forecasting technique performs way better than the other time series techniques for this type of data.

VI. CONCLUSIONS AND FUTURE WORK

Conclusion

This research analyses the various sales forecasting techniques used to predict sales for a tourism firm. Here the performance of Holt’s Winter method is compared with the performance of other time series technique. The result shows that Holt’s winter method performs better than the other time series forecasting techniques.

Sentiment Analysis is an important area of investigation. As the Web applications produce enormously and collect meaningful information, mining such information has become an important task. This research explores the predictive power of reviews using the tourism domain as a case study, and predicts sales using sentiment information mined from reviews. The sales prediction is calculated as a combination of sentiment analysis score and Holt’s winter method. The derived result states that the forecasted value is varied based on the sentiment analysis score for past months.
Future Work
For future work, the predictive power of reviews to predict customer behavior and utilize it for improving the company’s business can be implemented. Also different algorithms like ARSA can be applied on the current dataset and their accuracy in forecasting can be calculated.

REFERENCES