

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:

Section II reviews the literature survey. This gives a brief overview of Sentiment Analysis, Sales Forecasting Techniques, Methodologies and Parameters related to Sales Forecasting.

Section III is devoted to the design methodology. It presents the proposed sales forecasting architecture, flow of the Project with different UML diagrams and UI design.

Section IV presents implementation algorithms of sentiment analysis and sales forecasting.

Section V discloses Results and Analysis of results. It describes the effect of seasons on sales prediction. Also it describes the comparison of different time series forecasting techniques.

Section VI presents conclusion and scope for future research. It explains in details the major findings that are concluded in the research.

The main objective of this research is to Analyze various Forecasting techniques, understand various Sentiment Analysis algorithms and develop a system that analyses the reviews of Angel Travels, derive sentiments out of it and predicts sales based on the past sales data.

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.

Gautami 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 Beheshti-Kashia et.al. [8] presented methods in the sales forecasting research with a focus on fashion and new product forecasting.

III. PROPOSED SYSTEM

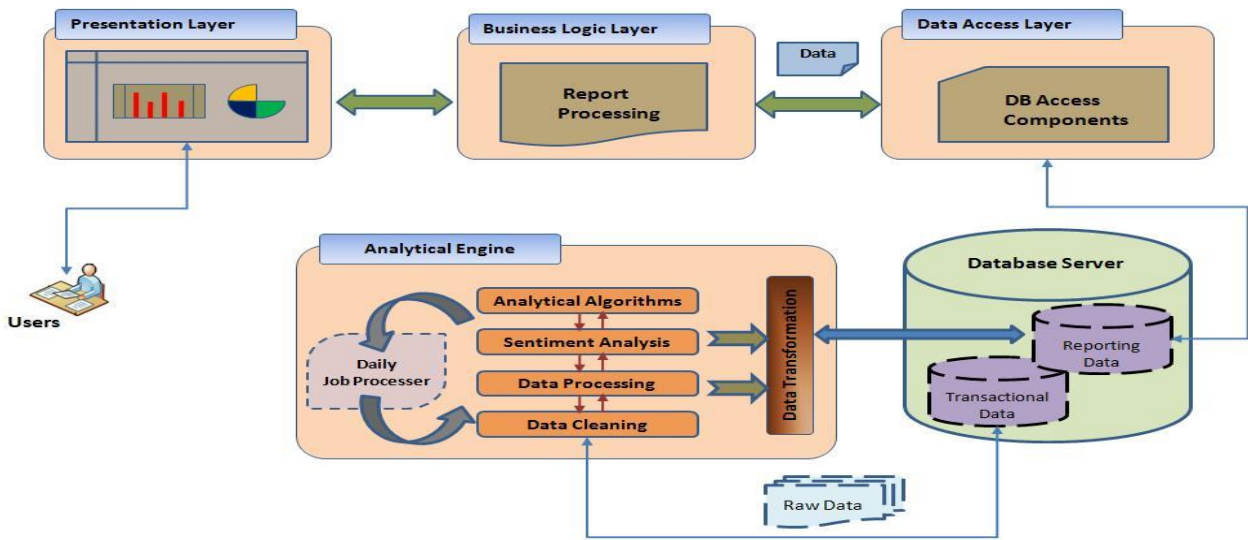


Figure 3. Block diagram for proposed system

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

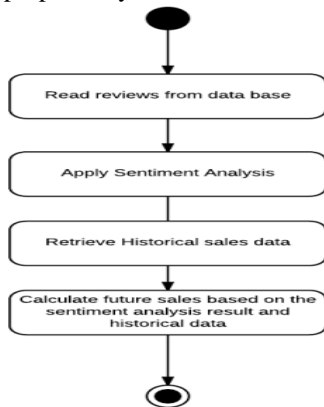


Figure 3.1. System flow diagram

3.2 UML Diagrams

Use Case Diagram

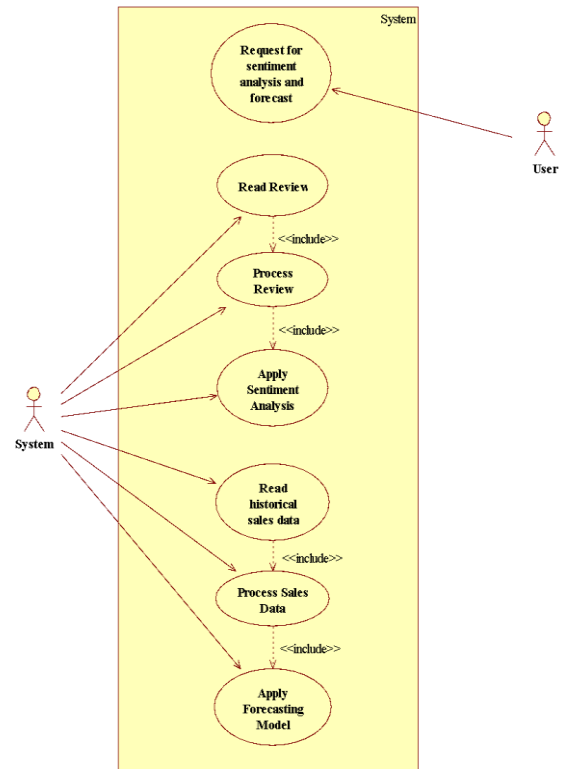


Figure 3.2.1 Use Case Diagram

Sequence Diagram

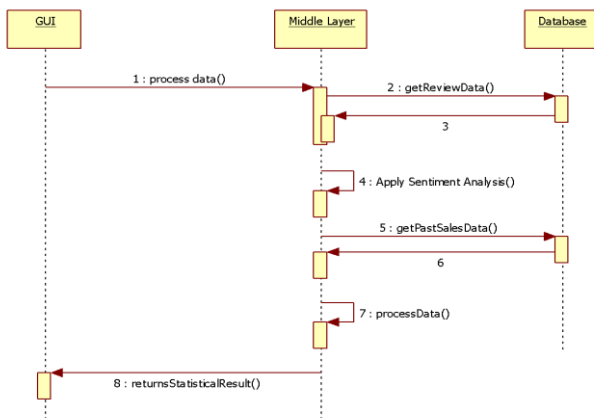


Figure 3.2.2 Sequence Diagram of Sales Prediction

1. Select the tour for which the prediction has to be done and start the process.
2. Read review data from the database.
3. Review data will be returned to the middle tier.
4. Apply sentiment analysis algorithm on the reviews and calculate average rating.
5. Get past sales data from the database.
6. Past sales data will be returned to the middle tier.
7. Apply algorithm for sales forecasting influenced by sentiment analysis on the sales data.
8. Return sales forecasting results to the GUI.

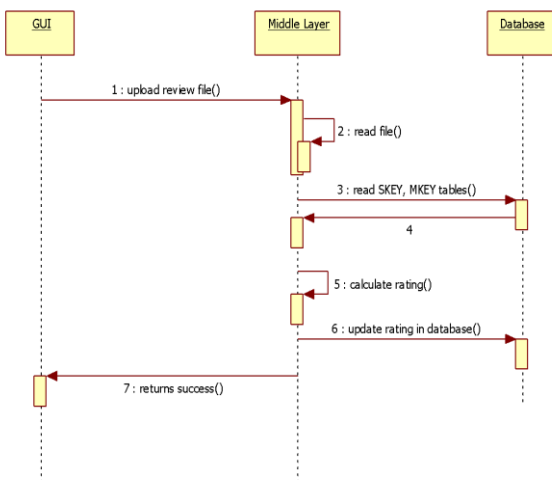


Figure 3.3.3 Sequence Diagram of Sentiment Analysis

1. Upload an excel file containing reviews for a specified tour.
2. Read the uploaded file.
3. Get single word keys (SKEY) and multi word keys (MKEY) from the database.
4. SKEY and MKEY data will be returned to the middle tier.
5. Compare the words with the reviews and calculate rating.
6. Update the rating in the database.
7. Success result will be sent to the GUI.

3.3 UI Design

Tourism Company Website Home Page

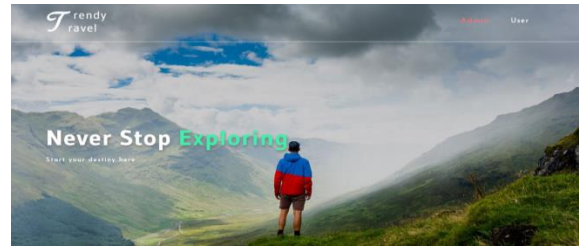


Figure 3.3.1 Home Page of Trendy Travel

Login Screen for User and Admin

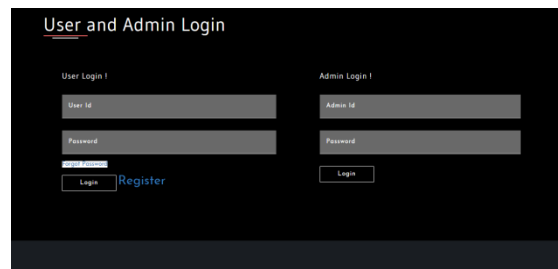


Figure 3.3.2 User and Admin Login Form

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)

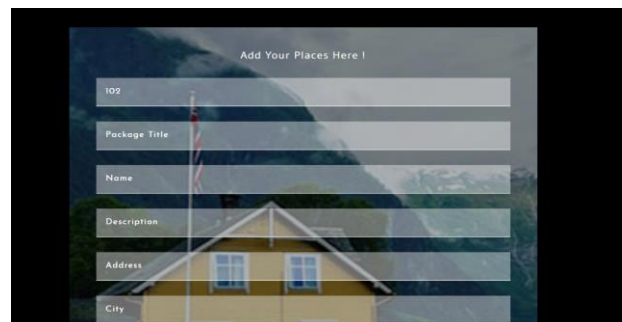


Figure 3.3.3 Add Places Form for Admin

System provides facility to upload Review files. Accessible only to Admin

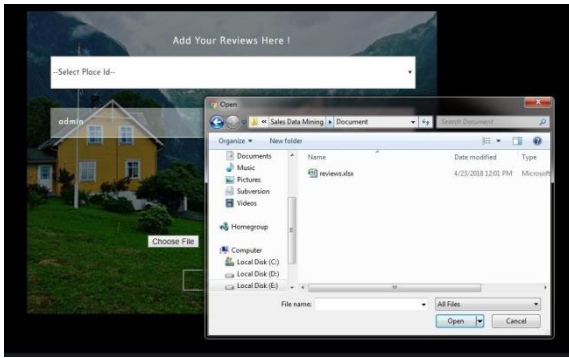


Figure 3.3.4 Add Reviews Form for Admin

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

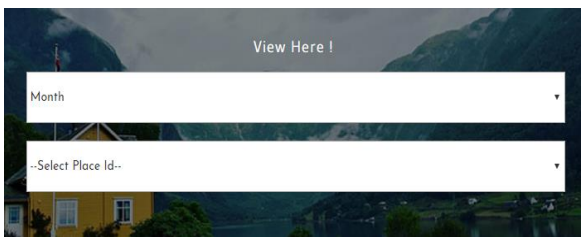


Figure 3.3.5 View Reviews and Rating form for Admin

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

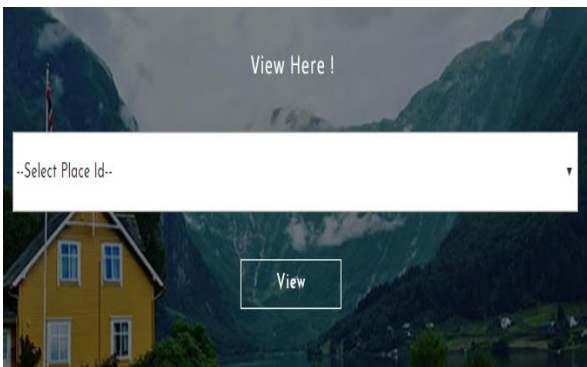


Figure 3.3.6 View Sales Report form for Admin

User Module Description

Register: User can register themselves and into the system using basic details.

Login: User can Login into the system, using his email id and password.

Forgot Password: If an user forgets his password, he can press forgot password and password will be sent to his E-Mail.

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

Send Feedback: User can send feedback to admin mentioning any issues they face, or any new places they want to add etc.

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



Figure 3.3.7 Tour Package form for User

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

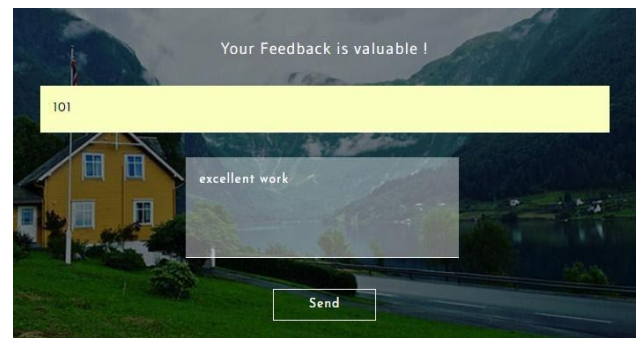


Figure 3.3.8 Feedback form for User

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.

1	Date	comments	rating
2	24/04/2017	Good Place	7
3	24/04/2017	Average place	5
4	24/04/2017	Very beautiful	8
5	24/04/2017	Not Bad for Summer	6
6	24/04/2017	Very beautiful	9
7	27/04/2017	Not Bad for Summer	4
8	27/04/2017	best place for winter holiday	8
9	27/04/2017	best place for Summer holiday	8
10	27/04/2017	best place for fall holiday	8
11	27/04/2017	great hospitality. Worth every penny	10
12	27/04/2017	brilliant location for photography	9
13	27/04/2017	summer is very harsh, the place is very scenic	7
14	27/04/2017	not a good place, not recommended	4
15	28/05/2017	lot of water logging not recommended during rainy season	3
16	28/05/2017	no electricity for most of the time, good place for relaxing though	7
17	28/05/2017	hotels are very poor and so are the people around , not recommended for families	5
18	28/05/2017	hotels are very nice, home stay is recommended, must go for lone travelers	8
19	28/05/2017	nice and relaxing place, best during summers	9
20	28/05/2017	clean and nice place, best to travel with families	7
21	30/05/2017	the connectivity of the place is very bad, reaching the place is very difficult	6
22	30/05/2017	located in a very nice and easily accessible place, very spacious and beautiful	8
23	30/05/2017	best place to spend the weekend	9
24	30/05/2017	very budget friendly, and nice and relaxing	8

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

Comments	Rating
Good Place	7
Average place	5
Very beautiful	8
Very beautiful	9
Not Bad for Summer	4
good	7
great	9
good	7
excellent	10
bad	2

Figure 5.2 Sentiment Analysis score

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

	A	B	C
1	Date	TourID	Tour
2	8/15/2017	101	Singapore Holiday
3	8/15/2017	101	Singapore Holiday
4	8/15/2017	101	Singapore Holiday
5	8/15/2017	101	Singapore Holiday
6	8/15/2017	101	Singapore Holiday
7	8/15/2017	101	Singapore Holiday
8	8/15/2017	101	Singapore Holiday
9	8/15/2017	101	Singapore Holiday
10	9/15/2017	101	Singapore Holiday
11	9/15/2017	101	Singapore Holiday
12	9/15/2017	102	Switzerland Holiday
13	9/15/2017	102	Switzerland Holiday
14	9/15/2017	102	Switzerland Holiday
15	9/15/2017	102	Switzerland Holiday
16	9/15/2017	102	Switzerland Holiday
17	9/15/2017	102	Switzerland Holiday
18	9/15/2017	102	Switzerland Holiday
19	9/15/2017	102	Switzerland Holiday
20	9/15/2017	102	Switzerland Holiday
21	9/15/2017	102	Switzerland Holiday
22	7/15/2017	102	Switzerland Holiday
23	7/15/2017	102	Switzerland Holiday
24	7/15/2017	102	Switzerland Holiday

Figure 5.3 Sales data for Tourism industry

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

Month	Sales Value	Forecast	Error	Absolute Error	Percent Error	Absolute Percent Error
MAY 2017	35	35.00	0.00	0.00	0.00	0.00
JUN 2017	30	35.00	5.00	5.00	0.17	0.17
JUL 2017	27	34.00	7.00	7.00	0.26	0.26
AUG 2017	28	32.60	4.60	4.60	0.16	0.16
SEP 2017	30	31.68	1.68	1.68	0.06	0.06
OCT 2017	34	30.34	-3.66	3.66	-0.11	0.11
NOV 2017	35	31.70	-3.30	3.30	-0.09	0.09
DEC 2017	40	33.72	-6.28	6.28	-0.16	0.16
JAN 2018	38	38.59	0.59	0.59	0.02	0.02
FEB 2018	30	38.17	8.17	8.17	0.27	0.27
MAR 2018	30	36.54	6.54	6.54	0.22	0.22
APR 2018	33	35.23	2.23	2.23	0.07	0.07
MAY 2018		35.23				

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

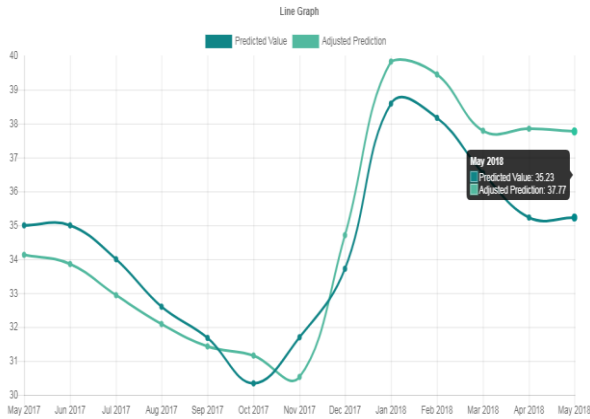


Figure 5.5 Sales forecast influenced by Sentiment Analysis

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

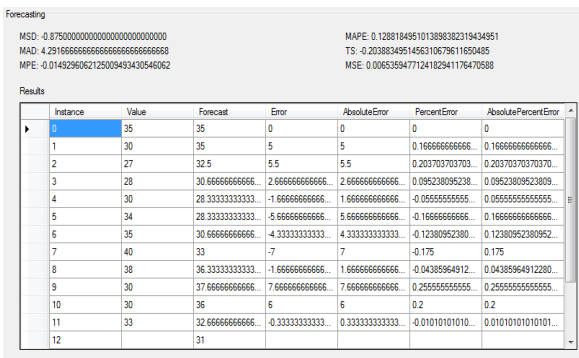


Figure 5.6 Simple Moving Average Method Forecast

Forecasting using Weighted Moving Average

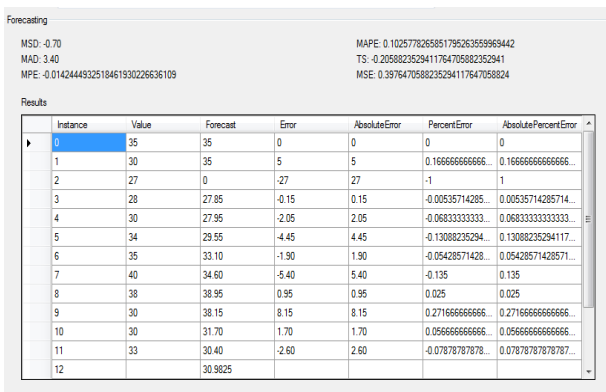


Figure 5.7 Weighted Moving Average Method Forecast

Forecasting using Exponential Smoothing

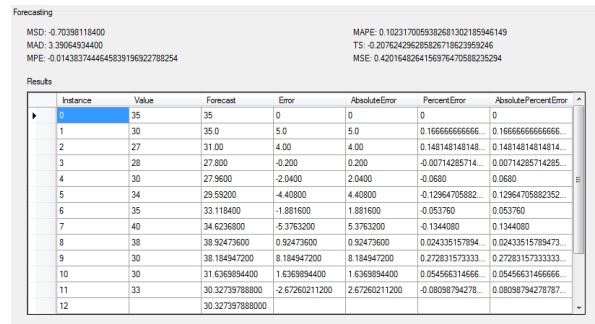


Figure 5.8 Exponential Smoothing Method Forecast

Forecasting using Adaptive Rate Smoothing/Holt's Winters Forecasting

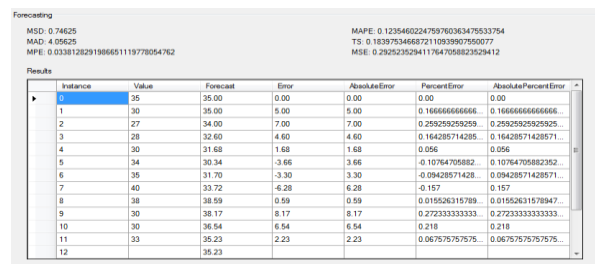


Figure 5.9 Holt's Winters Method Forecast

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

- [1] Beltran Borja Fiz Pontiveros, "Opinion Mining from Large Corpora of Natural Language Reviews", September 2012.
- [2] Richa Sharma, Shweta Nigam and Rekha Jain, "Opinion Mining Of Movie Reviews At Document Level", *International Journal on Information Theory (IJIT)*, Vol.3, No.3, July 2014.
- [3] A. Nisha Jebaseeli and E. Kirubakaran, "A Survey on Sentiment Analysis of (Product) Reviews", *International Journal of Computer Applications* (0975 – 888) Volume 47– No.11, June 2012.
- [4] Gautami Tripathi and Naganna S, "Opinion Mining: A Review", *International Journal of Information & Computation Technology*, ISSN 0974-2239, Volume 4, Number 16 2014, pp. 1625-1635, 2014.
- [5] Gurudeo Anand Tularam, Tareq Saeed, "Oil-Price Forecasting Based on Various Univariate Time-Series Models", 18 May 2016.
- [6] Prajakta S. Kalekar, "Time series Forecasting using Holt-Winters Exponential Smoothing", December 6, 2004.
- [7] Rui Yao and Jianhua Chen, "Predicting Movie Sales Revenue using Online Reviews", *IEEE International Conference on Granular Computing (GrC)*, 2013.
- [8] Samaneh Beheshti-Kashia, Hamid Reza Karimic, Klaus-Dieter Thoben, Michael Lütjen and Michael Teucke, "A survey on retail sales forecasting and prediction in fashion markets", *Systems Science & Control Engineering: An Open Access Journal*, Vol. 3, December 2014.
- [9] Patil Makrand Anil, "Sentiment Analysis", 2014.
- [10] Ravendra Ratan Singh Jandail, "A proposed Novel Approach for Sentiment Analysis and Opinion Mining", *International Journal of UbiComp (IJU)*, Vol.5, No.1/2, April 2014.
- [11] wikipedia.org/wiki/MouthShut.com accessed on 20th July 2016.
- [12] <http://www.mouthshut.com> accessed on 20th July 2016.
- [13] <http://www.mouthshut.com/websites/MakeMyTrip-com-reviews> accessed on 20th July 2016.
- [14] Nihalahmad R. Shikalgar, Deepak Badgujar, "Online Review Mining for Forecasting Sales", *IJRET: International Journal of Research in Engineering and Technology*, eISSN: 2319-1163 | pISSN: 2321-7308, Volume: 02, Dec-2013.
- [15] Viet Hung Nguyen & Zhuochuan Wang, "Practice of Online Marketing with Social Media in Tourism Destination Marketing: The case study of VisitSweden", 2011.
- [16] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs," *Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, pp. 607-614, 2007.
- [17] Rui Yao and Jianhua Chen, "Predicting Movie Sales Revenue using Online Reviews", *IEEE International Conference on Granular Computing (GrC)*, 2013.
- [18] Xiaohui Yuy, Yang Liu and Aijun An, "An Adaptive Model for Probabilistic Sentiment Analysis", *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2010.
- [19] Fabian Abel, Ernesto Diaz-Aviles, Nicola Henze, Daniel Krause and Patrick Siehdnel, "Analyzing the Blogosphere for Predicting the Success of Music and Movie Products" *IEEE International Conference on Advances in Social Networks Analysis and Mining*, 2010.
- [20] A. Nisha Jebaseeli and E. Kirubakaran, "A Survey on Sentiment Analysis of (Product) Reviews", *International Journal of Computer Applications* (0975 – 888) Volume 47– No.11, June 2012.
- [21] Yi-Ching Zeng and Shih-Hung Wu, "Modeling the Helpful Opinion Mining of Online Consumer Reviews as a Classification Problem", *IJCNLP Workshop on Natural Language Processing for Social Media (SocialNLP)*, October 2013.
- [22] Arti Buche, Dr. M. B. Chandak and Akshay Zadgaonkar, "Opinion Mining And Analysis: A Survey", *International Journal on Natural Language Computing (IJNLC)*, Vol. 2, No.3, June 2013.
- [23] Gautami Tripathi and Naganna S, "Opinion Mining: A Review", *International Journal of Information & Computation Technology*, ISSN 0974-2239, Volume 4, Number 16 2014, pp. 1625-1635, 2014.
- [24] Arzu Baloglu and Mehmet S. Aktas, "BlogMiner: Web Blog Mining Application for Classification of Movie Reviews", *IEEE Fifth International Conference on Internet and Web Applications and Services*, 2010.
- [25] Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, and Kang Zhang, "SentiView: Sentiment Analysis and Visualization for Internet Popular Topics", *IEEE Transactions on Human-Machine Systems*, Volume: 43, No. 6, November 2013.