End Project Report on

"SALES PREDICTION USING MARKETING SPENDS"

Submitted for end project of Introduction to Artificial Intelligence and Machine Learning by:

Swarup Shinde   22ECB0A13

Supervised By:

- Sri. Ravi Kishore Kodali

Department of Electronics and

Communication Engineering

National Institute of Technology, Warangal

December, 2023

TITLE**:**

SALES PREDICTION USING MARKETING SPENDS

Abstract:

This study explores the correlation between marketing expenditures and sales predictions, employing advanced analytical models to uncover patterns and trends. By leveraging historical data on marketing spends and corresponding sales outcomes, the research aims to develop a predictive framework for sales performance. The analysis involves various machine learning techniques to discern the impact of different marketing channels and allocation strategies on future sales, providing valuable insights for optimizing marketing budgets and enhancing overall business forecasting accuracy.

Introduction:

In the dynamic landscape of modern business, the effective allocation of resources, particularly in marketing endeavours, is pivotal for sustainable growth. This study delves into the intricate relationship between marketing expenditures and sales predictions, recognizing the pressing need for businesses to optimize their strategies. By harnessing the power of advanced analytics and machine learning, this research seeks to unravel the nuanced patterns within historical data. Through a comprehensive examination of marketing spends and corresponding sales outcomes,

the aim is to construct a predictive framework that empowers organizations to make informed decisions.

What is Simple Linear Regression?

This study explores the foundational concept of simple linear regression as a powerful statistical tool for modeling the relationship between two variables. Focusing on the fundamental principles of this regression technique, the research aims to elucidate its application in predicting and understanding the behavior of a dependent variable based on a single independent variable. Through a systematic analysis of real-world data, the study highlights the simplicity and interpretability of simple linear regression, making it an accessible yet robust method for exploring linear relationships.

Related Work:

The related work in field of sales prediction is also done by using the data of video game developing companies:

- There are a lot of resources on the internet about finding insights and training models on machine learning datasets however very few articles on how to use these models for building actual applications.

- So today we are going to learn this process by first training a video game sales prediction model using a dataset from a hackathon and then use the trained model for creating a basic app that gives us sales prediction based on user inputs.

- This article is divided into sections that you can pick up one by one instead of trying to finish it one go. It took me a full week to finish the app from the point when I first picked up the dataset. Therefore, take your own time and focus on learning various aspects of building the app rather than the final product.

Reference:

From Internet, website: Sales Prediction Using Machine Learning (analyticsvidhya.com)

My work:

I have done the implementation of the workflow in the KNIME Analytics Platform.
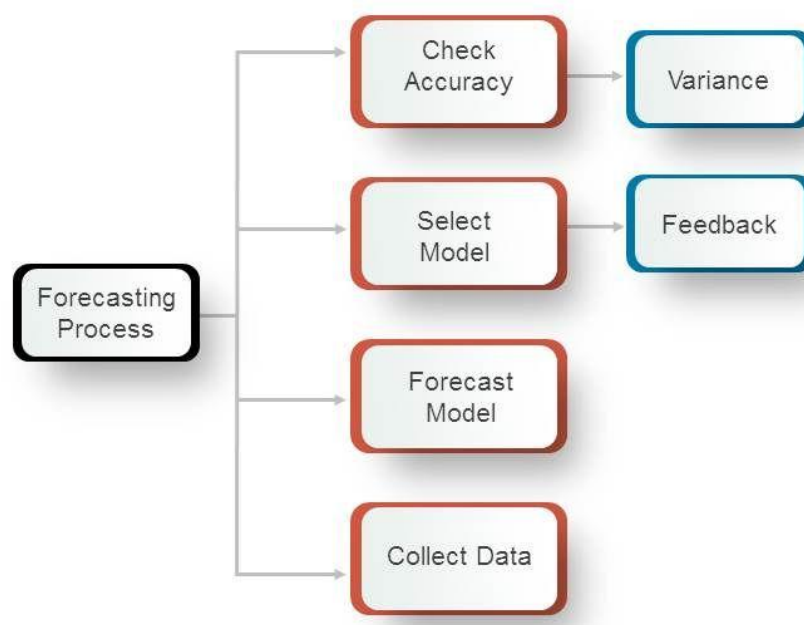

SALES PREDICTION FLOWCHART MODEL

About the project:

The project uses simple regression learning model for the prediction of sales using marketing spends. The workflow shows how to use simple regression learner to predict the value of a numerical target column (sales).

Software Used: KNIME Analytics Platform

Predicting sales using marketing spends is a data frame containing the impact of three medias (TV, Radio and Newspaper) on sales. Data are the advertising budget in thousands of dollars along with the sales. The advertising experiments has been repeated 200 times.

Implementation of Workflow in KNIME:
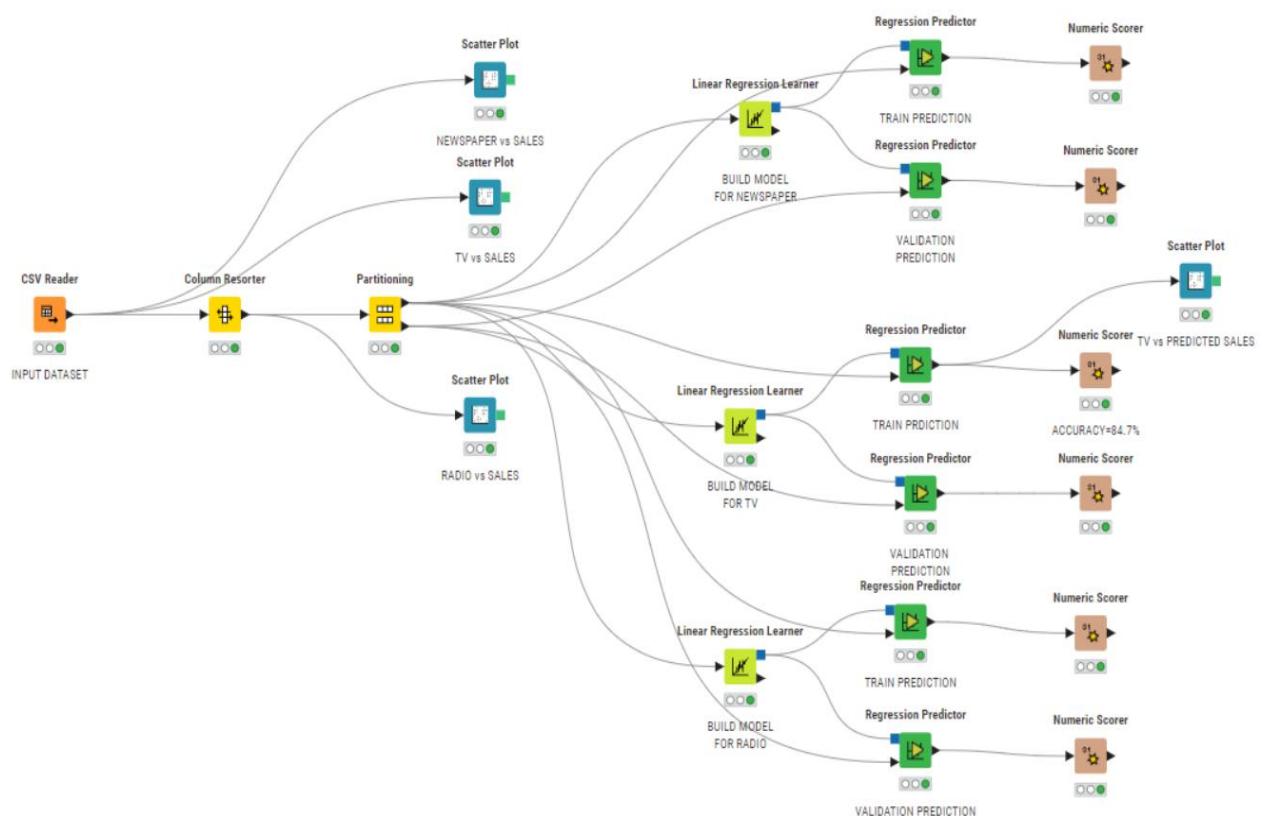
About the Dataset:

Source: I have collected the dataset from the open-source website Kaggle:

Website: Sales Prediction (Simple Linear Regression) | Kaggle

By using the dataset we can understand the relationship between "sales", "tv", "radio" and "newspaper" using simple linear regression.

This dataset contains of 4 columns and around total of 10k rows of data in it. Predicting sales using marketing spends is a data frame containing the impact of three medias (TV, Radio and Newspaper) on sales. Data are the advertising budget in thousands of dollars along with the sales. The advertising experiments has been repeated 200 times.

Dataset:

# TV	# Radio	# Newspaper	# Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2
8.6	2.1	1	4.8
199.8	2.6	21.2	15.6
66.1	5.8	24.2	12.6
214.7	24	4	17.4
23.8	35.1	65.9	9.2
97.5	7.6	7.2	13.7
204.1	32.9	46	19
195.4	47.7	52.9	22.4
67.8	36.6	114	12.5
281.4	39.6	55.8	24.4
69.2	20.5	18.3	11.3
147.3	23.9	19.1	14.6

Results:

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
TV	0.0541	0.0022	24.1287	0.0
Intercept	7.2268	0.388	18.6246	0.0

R-Squared: 0.7865
Adjusted R-Squared: 0.7852

1. **Output of linear regression learner**

R^2:	0.787
Mean absolute error:	1.922
Mean squared error:	5.624
Root mean squared error:	2.371
Mean signed difference:	0
Mean absolute percentage error:	0.153
Adjusted R^2:	0.787

2. Statistics

Accuracy:
=(1-mean absolute   percentage error) x 100
=(1-0.153) x 100
=0.847 x 100
= 84.7%

Accuracy = 84.7%

Table View

Rows: 160 | Columns: 5

Row...	Sales Number (double)	TV Number (double)	Radio Number (double)	Newspaper Number (double)	Prediction (Sales) Number (double)
Row1	10.4	44.5	39.3	45.1	9.633
Row2	12	17.2	45.9	69.3	8.157
Row3	16.5	151.5	41.3	58.5	15.418
Row4	17.9	180.8	10.8	58.4	17.002
Row5	7.2	8.7	48.9	75	7.697
Row6	11.8	57.5	32.8	23.5	10.336
Row7	13.2	120.2	19.6	11.6	13.726
Row8	4.8	8.6	2.1	1	7.692
Row9	15.6	199.8	2.6	21.2	18.029
Row...	12.6	66.1	5.8	24.2	10.801
Row...	17.4	214.7	24	4	18.835
Row...	13.7	97.5	7.6	7.2	12.498
Row...	22.4	195.4	47.7	52.9	17.792
Row...	12.5	67.8	36.6	114	10.893
Row...	11.3	69.2	20.5	18.3	10.968
Row...	17.5	237.4	5.1	23.5	20.062
Row...	5.6	13.2	15.9	49.6	7.94
Row...	20.5	228.3	16.9	26.2	19.57
Row...	9.7	62.3	12.6	18.3	10.595

3. Predicted Sales result from linear regression predictor

Name	Type	# Missing v...	# Unique val...	Minimum	Maximum	25% Quantile	50% Quantil...	75% Quantile	Mean	Mean Absol...
Sales	Number (dou...	0	105	1.6	27	11.3	16.05	19.55	15.408	4.282
TV	Number (dou...	0	153	0.7	296.4	76.325	154.05	221.925	151.317	73.773
Radio	Number (dou...	0	139	0	49.6	9.675	21.2	36.75	23.027	13.211
Newspaper	Number (dou...	0	138	0.3	114	12.45	24.95	43.275	30.114	17.486
Prediction (Sa...	Number (dou...	0	153	7.265	23.252	11.353	15.556	19.226	15.408	3.989

4. Statistics of Predictor

Conclusion:

At present, the sales prediction is a necessary process for all the industries to develop In the future. In this project, there are many different ways to implement this process using other machine learning models. I have completed my required work based on the implementation of sales prediction using marketing spends. The accuracy of the machine learning model is about 85%. So, I conclude that the model created is accurate enough to predict sales by using linear regression machine learning model.

Reference:

1. Internet (Related Work):

   Website: Sales Prediction Using Machine Learning (analyticsvidhya.com)

2. Kaggle (Dataset):

   Website: Sales Prediction (Simple Linear Regression) | Kaggle

3. KNIME website (other information):

   Website: Open for Innovation | KNIME