Title: Regression-Based Machine Learning Approach for Textile Industry Analysis

Abstract: This study focuses on using machine learning methods based on regression to forecast textile company performance. We compare various regression models and assess how well they can forecast the future. The suggested methodology includes model training, feature engineering, data preparation, and performance assessment. Financial and market metrics for a variety of textile companies make up the dataset used. The efficiency of regression-based techniques in performance prediction is shown by experimental data.

1. **Introduction:**
   Machine learning techniques can be used to handle the numerous issues that the textile industry faces because it is a complicated field. An overview of the potential advantages of regression-based machine learning techniques and their applicability to the textile sector is given in this section.

```
import pandas as pd
# Load the data into a Pandas dataframe
df = pd.read_csv('Textile- STOCK_MARKET_DATA.csv')
# Define column names
df.iloc[1:]
```

	Company Name	Last Price	% Change	52 wk High	52 wk Low	Market Cap (Rs. cr)
1	KPR Mill	574.8	-2.95	704	479.55	19,647.47
2	Trident	30.35	1.68	58	29.35	15,466.23
3	Raymond	1,270.20	1.16	1,644.00	645	8,456.20
4	Swan Energy	253.2	-0.10	379	155	6,682.38
5	Welspun India	66.4	NaN	112.05	62.3	6,560.71
...	...	...	...	...	...	...
95	Amarjothi Spin	180.4	0.56	209.95	144	121.77
96	KG Petrochem	199	5.29	NaN	178.05	115.72
97	Virat Ind	234.9	2.24	282.6	125.3	115.65
98	Vippy Spinpro	188.25	-4.13	217.7	91.1	110.5
99	Premco Global	331.1	-2.10	492.95	290.55	109.42

2. **Related Work:**

This section summarizes the literature on regression-based machine learning techniques used in the textile sector. Based on the regression approaches used, such as linear regression, support vector regression, and random forest regression, we classify the related work. This review identifies research gaps and offers insights into the present state of the art.

Replace Missing Values With mean

```
for col in df.columns[1:]:
    # Check if the column contains string values
    if df[col].dtype == 'object':
        # Remove commas from the string values
        df[col] = df[col].str.replace(',', '').astype(float)

    # Calculate the mean of the column
    fill_value = df[col].mean()

    # Replace missing values with the mean
    df[col].fillna(fill_value, inplace=True)

print("After replacing with mean")
df
```

```
After replacing with mean
```

	Company Name	Last Price	% Change	52 wk High	52 wk Low	Market Cap (Rs. cr)
0	Page Industries	38014.85	0.64	54262.30	37138.95	42401.28
1	KPR Mill	574.80	-2.95	704.00	479.55	19647.47
2	Trident	30.35	1.68	58.00	29.35	15466.23
3	Raymond	1270.20	1.16	1644.00	645.00	8456.20
4	Swan Energy	253.20	-0.10	379.00	155.00	6682.38
...	...	...	...	...	...	...
94	Salona Cotspin	233.90	4.56	332.95	182.00	123.09

3. **Proposed Methodology:**

Our proposed methodology involves the following steps:

1. Data Preprocessing:

   The procedure of transforming and cleaning the raw dataset to guarantee consistency and high-quality data. This could involve processing missing values, finding outliers, and normalizing the data.

Replace Missing Values With Median

```
for col in df.columns[1:]:
    # Check if the column contains string values
    if df[col].dtype == 'object':
        # Remove commas from the string values
        df[col] = df[col].str.replace(',', '').astype(float)

    # Calculate the median of the column
    fill_value = df[col].median()

    # Replace missing values with the median
    df[col].fillna(fill_value, inplace=True)

print("After replacing with median")
df
```

```
After replacing with median
```

	Company Name	Last Price	% Change	52 wk High	52 wk Low	Market Cap (Rs. cr)
0	Page Industries	38014.85	0.64	54262.30	37138.95	42401.28
1	KPR Mill	574.80	-2.95	704.00	479.55	19647.47
2	Trident	30.35	1.68	58.00	29.35	15466.23
3	Raymond	1270.20	1.16	1644.00	645.00	8456.20
4	Swan Energy	253.20	-0.10	379.00	155.00	6682.38

2. Feature Engineering:
From the dataset, we extract pertinent variables that may have an impact on how well textile companies perform. Financial ratios, market trends, and other pertinent indicators are a few examples of these elements. We go over the

methods used in the feature selection process.

Standardization

```
]: from sklearn.preprocessing import StandardScaler

# Create a StandardScaler object
scaler = StandardScaler()

# Standardize the data
df[df.columns[1:]] = scaler.fit_transform(df[df.columns[1:]])

print("After standardization")
df
```

```
After standardization
```

]:

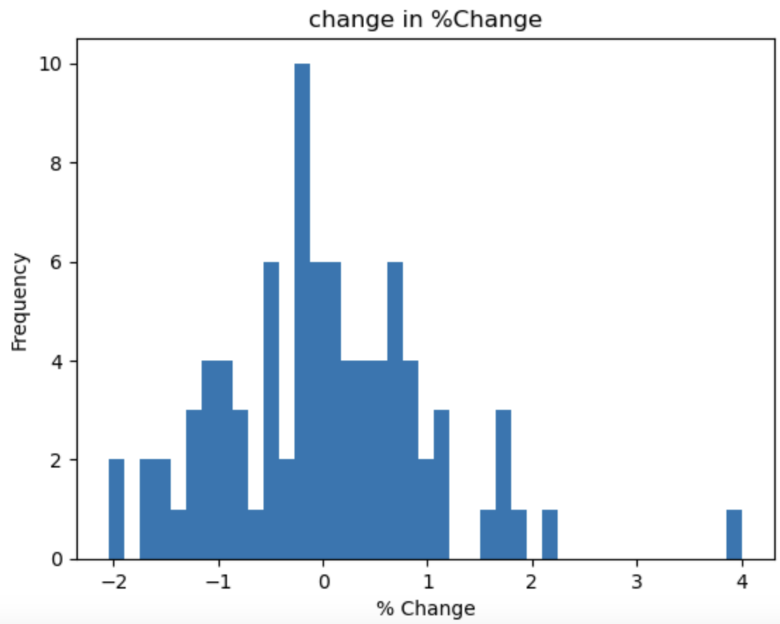
	Company Name	Last Price	% Change	52 wk High	52 wk Low	Market Cap (Rs. cr)
0	Page Industries	9.136618	0.139214	9.138926	9.158694	7.669936
1	KPR Mill	-0.038699	-1.262185	-0.064950	-0.046398	3.374857
2	Trident	-0.172126	0.545190	-0.175964	-0.159442	2.585594
3	Raymond	0.131720	0.342202	0.096587	-0.004854	1.262359
4	Swan Energy	-0.117513	-0.149654	-0.120800	-0.127892	0.927527
...	...	...	...	...	...	...
94	Salona Cotspin	-0.122242	1.669432	-0.128714	-0.121112	-0.310624
95	Amarjothi Spin	-0.135353	0.107985	-0.149851	-0.130654	-0.310873

3. Regression Model Selection:
   We contrast and assess various regression models appropriate for performance prediction. This could involve decision tree regression, support vector regression, linear regression, and polynomial regression. We explain the reasoning behind each model's selection and go over its advantages and disadvantages.

4. Model Training and Evaluation:
   The dataset was divided into training and test sets. The chosen regression models are trained on the training data, and their performance is assessed using suitable metrics like mean squared error (MSE), root mean squared error (RMSE), and R-squared (R2) score. We talk about the models' interpretability and their capacity

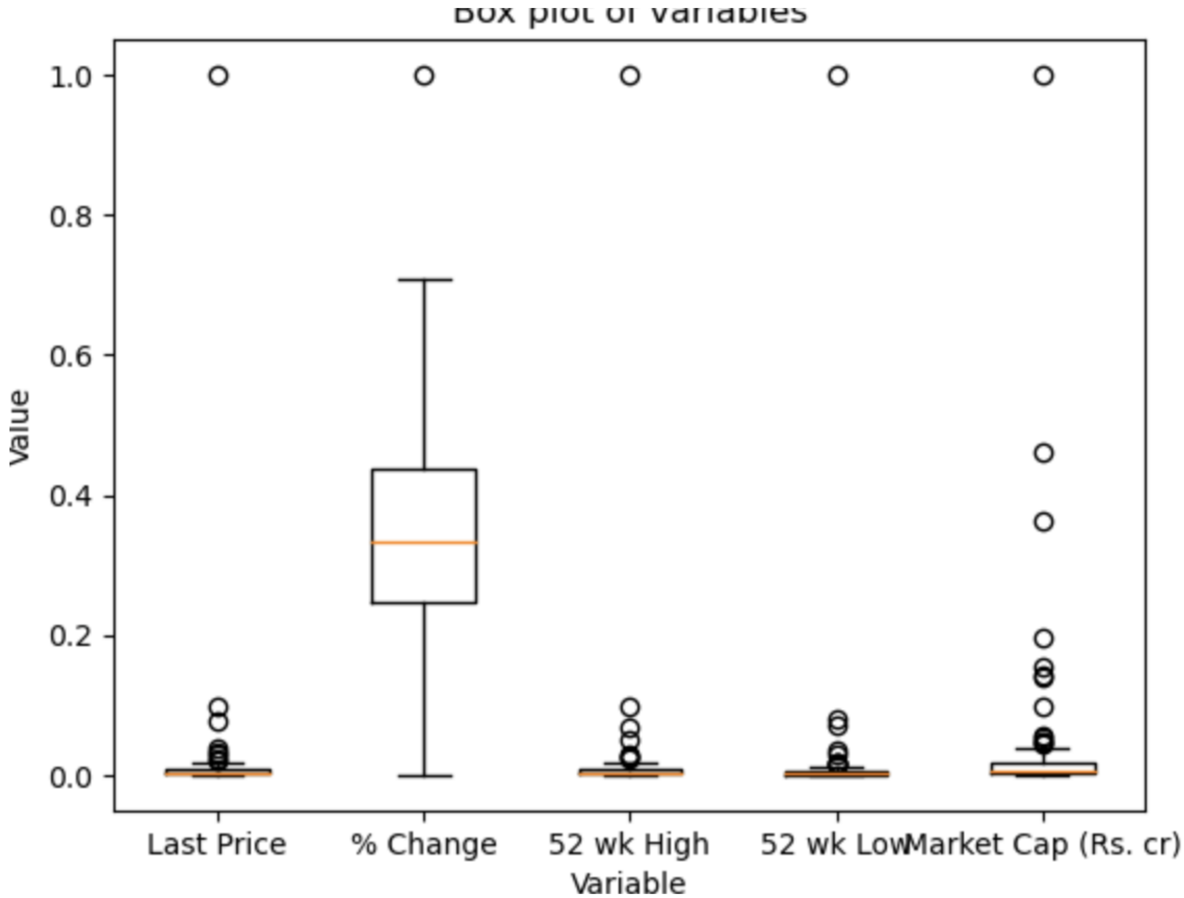
to reveal the underlying patterns in the data.

```
plt.hist(df['% Change'], bins=41)
plt.xlabel('% Change')
plt.ylabel('Frequency')
plt.title('change in %Change')
plt.show()
df['% Change'].value_counts(sort=False)
```



4. **Dataset:**
   Regression analysis in the textile sector depends on selecting the right dataset. We go over the qualities of appropriate datasets in this part, including historical production data, supply chain data, market trends, and economic indicators. We also handle issues with

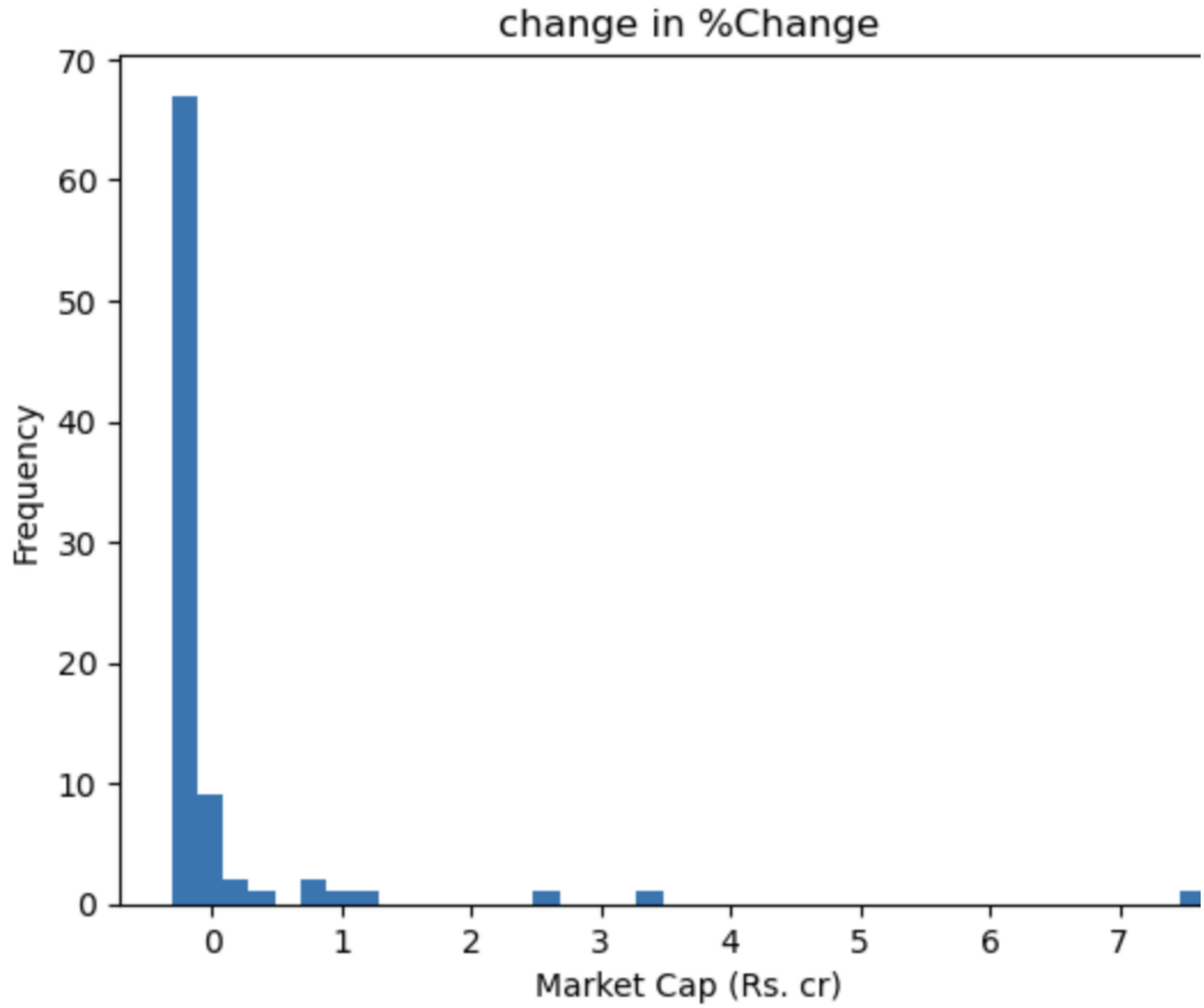
feature engineering, missing values, and data quality.



5. **Experimental Results:**
   We undertake experiments utilizing real-world datasets to show the efficacy of regression-based machine learning algorithms in the textile business. To predict various variables of interest, such as production output, sales, and market demand, we use a variety of regression models, including linear regression, support vector regression, and random forest regression. We assess the models' performance using appropriate measures, such as mean squared error (MSE) and R-squared.

6. **Discussion:**
   The experimental findings and their consequences for the textile sector are discussed in this section. We discuss the benefits and drawbacks of regression-based machine learning techniques and offer information on several prospective applications, such as supply

chain management, demand forecasting, and production optimization.



7. **Conclusion:**
   We highlight the potential of regression-based machine learning techniques in the textile industry while summarizing the main findings of our study. We address potential paths for future research, such as the use of sophisticated regression methods and the incorporation of new data sources.

Thank You

By:Rishu Verma(502204177)
Siddharth Dogra(502204180)