r value is a way to determine Pearson Correlation. The Pearson correlation coefficient's range is (-1, 1) [31]. When the value is positive, there is a positive linear correlation; when it is negative, there is a negative linear correlation. The linear association is stronger the closer the result is to +1 or 1 [32]. When the correlation coefficient reaches 1, it fully becomes positive. When the correlation coefficient is -1, it is fully negative. The correlation is stronger when the correlation coefficient's absolute value is higher. The association is weaker the closer the correlation coefficient is to 0 [33].

III. Results and Discussion

A. Data Visualization

First, the process started by downloading and uploading the .csv data to Google Drive. Second, opening the data in Google Spreadsheet and prepare several hypotheses to predict and determine whether a significant relationship exists between them. Researchers can use visualization tools to explore the relationship between variables from any angle and with any rotation to find relationships and to see how changes in the values of one outcome variable affect the values of other variables collectively [34].

The goal is to build visualizations and predictions to ascertain whether there is a substantial connection between several hypotheses:

- H1: Item_Weight has a significant relationship to Item_Outlet_Sales
- H2: Item_Visibility has a significant relationship with Item_Outlet_Sales
- H3: Item_MRP has a significant relationship with Item_Outlet_Sales

Third, the process continued by copying the relevant variables to test the hypotheses on different sheets. Fourth, creating a scatter plot for each hypothesis test. Figure 1 is represented the H1: Item_Weight has a significant relationship to Item_Outlet_Sales hypothesis test.

Fig. 1 Scatter plot for testing H1: Item_Weight has a significant relationship to Item_Outlet_Sales

A correlation was thought to be linear—that is, it should follow a line. According to the correlation theory, the H1 represented no correlation. Meaning that the values were not seem linked at all.

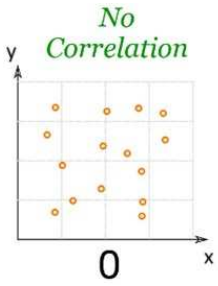

Fig. 2 H1 assumed correlation [35]

Fig. 3 Scatter plot for testing H2: Item_Visibility has a significant relationship with Item_Outlet_Sales

It was expected that a correlation would be linear—that is, follow a line. According to the correlation theory, the H2 represented low negative correlation. When one value rose while the other fell, there was a negative correlation.

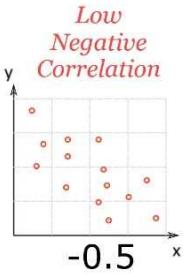

Fig. 4 H2 assumed correlation [35]

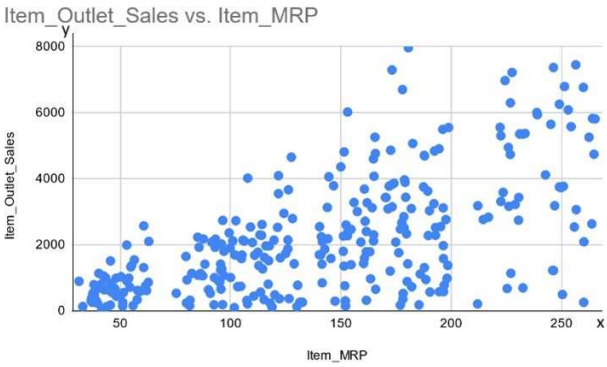

Fig. 5 Scatter plot for testing H3: Item_MRP has a significant relationship with Item_Outlet_Sales

It was expected that a correlation would be linear, or follow a line. According to the correlation theory, the H3 represented low positive correlation. When both values increased, there was a positive correlation.

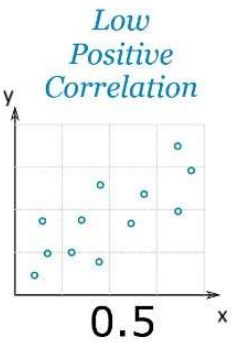

Fig. 6 H3 assumed correlation [35]

A. Pearson's Correlation and descriptive statistical analysis

The Pearson correlation test is an analytical technique utilized to decide whether exists a significant relationship among the two variables being tested. There are several steps to check Pearson's Correlation.

- *Step 1:* The next process is cleaning the data to check for the missing values.
- *Step 2:* Then calculate the mean or average value of each variable. The Pearson correlation test is an analytical technique used to determine whether a significant relationship exists between the two variables being tested. In addition to the correlation test, descriptive statistical analysis is also carried out here. With this simple data processing, the raw data will be more organized and easier to analyse, read, and use.
- *Step 3:* Subtract the mean of x from each value of x (designate them "a") and the mean of y from each value of y (designate them "b")
- *Step 4:* Determine: ab, a^2, and b^2 in each value.
- *Step 5:* Determine the total amount of ab, the total amount of a^2, and the total amount of b^2. *Step 6:* Divide the total of ab by the square root of [(total of a^2) × (total of b^2)]

An equation [36], it is:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

Where:

- $\sum$ is the "total" symbol, Sigma.
- $(x_i - \bar{x})$ is each x-value minus the mean of x (referred to as "a" above)
- $(y_i - \bar{y})$ is each y-value minus the mean of y (referred to as "b" above)

The H1 result is represented in Table 1. From the analysis conducted using the scatter plot method and the Pearson Correlation formula, we could see that the results obtained were 0.1 whereas the Pearson Correlation test results are quite close to 0. This condition showed no relationship between the two Item_Weight and Item_Outlet_Sales variables where the higher or lower Item_Weight then had no effect on Item_Outlet_Sales.

TABLE I
H1: ITEM_WEIGHT HAS A SIGNIFICANT RELATIONSHIP TO ITEM_OUTLET_SALES

Item_Weight Average score	Item_Outlet_Sales Average score	Pearson's Correlation
12.76123797	2156.076447	0.1148738262

The H2 result is represented in Table 2. From the analysis conducted using the scatter plot method and the Pearson Correlation formula, we can see that the results obtained were -0.071 whereas the results of the Pearson Correlation test were quite low which indicated that the correlation between the data was quite weak. From the scatterplot form, there was a strong but negative correlation where if Item_Visibility decreased, Item_Outlet_Sales will be increased and vice versa.

TABLE II
H2: ITEM_VISIBILITY HAS A SIGNIFICANT RELATIONSHIP WITH ITEM_OUTLET_SALES

Item_Weight Average score	Item_Outlet_Sales Average score	Pearson's Correlation
0.06846437962	2217.811643	-0.07387090645

The H3 result is represented in Table 3. From the analysis conducted using the scatter plot method and the Pearson Correlation formula, we can see that the results obtained were 0.6 whereas the Pearson Correlation test results were close to number 1. This condition showed that there was a positive relationship between the two Item_MRP and Item_Outlet_Sales variables, which was higher Item_MRP, the higher Item_Outlet_Sales.

TABLE III
H3: ITEM_MRP HAS A SIGNIFICANT RELATIONSHIP WITH ITEM_OUTLET_SALES

Item Weight Average score	Item_Outlet_Sales Average score	Pearson's Correlation
141.5765923	2223.950936	0.6211252581

Every business wants to be aware of customer demand in advance of any season to prevent product shortages. As time goes on, there will be an exponential rise in the need for businesses to make predictions with more accuracy. As a result, extensive study is being done in this field to make precise sales predictions. The company's profit is directly correlated with its ability to make better predictions. It has been attempted to predict sales in this study [11]. The correlation was examined using Pearson's correlation coefficient (r) methodology [32]. The research result shows there was no connection between the variables Item_Weight and Item_Outlet_Sales. Item_Outlet_Sales and Item_Visibility had a significant but inverse relationship when visibility falls. The variables Item_MRP and Item_Outlet_Sales were positively correlated.

IV. CONCLUSION

In a Big Mart Company, a sales forecast was applied to assess the availability of different items sold at different shops in different towns. As a result, it is important to determine how a product's attractiveness is influenced by various elements, including price, popularity, the time of day, the type of outlet, the location of the store, etc. According to the Pearson correlation test results, there was no connection between the variables Item_Weight and Item_Outlet_Sales. Item_Outlet_Sales and Item_Visibility had a significant but

inverse relationship when visibility falls. The variables Item_MRP and Item_Outlet_Sales were positively correlated.

The descriptive statistical analysis also supported the correlation test. The raw data will be more easily arranged, analyzed, read, and used after this simple data processing. The association between the several parameters assessed and the forecast results after implementation imply that additional stores might benefit. It is possible to create an effective recommendation system utilizing transactional data, which will enable customers with similar preferences to be recommended items from the business. To prevent unforeseen cash flow and to better manage production, labor, and financing requirements, it may be helpful to forecast sales and create a sales plan in advance.

ACKNOWLEDGMENT

We thank Research Institute and Community Service (LPPM) Universitas Pembangunan Nasional Veteran Jakarta, the Information Systems undergraduate program, Faculty of Computer Science Universitas Pembangunan Nasional Veteran Jakarta for the support.

REFERENCES

[1] T. Tjahjanto, A. Arista, and E. Ermatita, "Information System for State-owned inventories Management at the Faculty of Computer Science," Sinkron, vol. 7, no. 4, pp. 2182–2192, Oct. 2022, doi: 10.33395/sinkron.v7i4.11678.

[2] A. Arista and K. N. M. Ngafidin, "An Information System Risk Management of a Higher Education Computing Environment," International Journal on Advanced Science, Engineering and Information Technology, vol. 12, no. 2, p. 557, Apr. 2022, doi:10.18517/ijaseit.12.2.13953.

[3] A. Arista and B. S. Abbas, "Using the UTAUT2 model to explain teacher acceptance of work performance assessment system," International Journal of Evaluation and Research in Education (IJERE), vol. 11, no. 4, p. 2200, Dec. 2022, doi:10.11591/ijere.v11i4.22561.

[4] U. Rusdiana, I. Ernawati, N. Falih, and A. Arista, "Comparison of Distance Metrics on Fuzzy C-Means Algorithm Through Customer Segmentation," in *2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2021, pp. 307–311.

[5] W. Cholil, F. Panjaitan, F. Ferdiansyah, A. Arista, R. Astriratma, and T. Rahayu, "Comparison of Machine Learning Methods in Sentiment Analysis PeduliLindungi Applications," in *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, 2022, pp. 276–280.

[6] T. Theresiawati, H. B. Seta, and A. Arista, "Implementing quality function deployment using service quality and Kano model to the quality of e-learning," International Journal of Evaluation and Research in Education (IJERE), vol. 12, no. 3, p. 1560, Sep. 2023, doi:10.11591/ijere.v12i3.25511.

[7] N. Malik and K. Singh, "Sales Prediction Model for Big Mart," *Paricahy:Maharaja Surajmal Institute Journal of Applied Research*, vol. 3, no. 1, pp. 22–32, 2020.

[8] R. P and S. M, "Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), May 2021, doi:10.1109/iciccs51141.2021.9432109.

[9] M. K. Nishad and S. Kondekar, "BIG MART SALES PREDICTION," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 5, pp. 1698–1702, 2022, [Online]. Available: www.irjmets.com

[10] T. K. Thivakaran and M. Ramesh, "Exploratory Data analysis and sales forecasting of bigmart dataset using supervised and ANN algorithms," Measurement: Sensors, vol. 23, p. 100388, Oct. 2022, doi: 10.1016/j.measen.2022.100388.

[11] K. Punam, R. Pamula, and P. K. Jain, "A Two-Level Statistical Model for Big Mart Sales Prediction," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Sep. 2018, doi: 10.1109/gucon.2018.8675060.

[12] G. Behera and N. Nain, "Grid Search Optimization (GSO) Based Future Sales Prediction for Big Mart," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Nov. 2019, doi: 10.1109/sitis.2019.00038.

[13] G. Behera and N. Nain, "A Comparative Study of Big Mart Sales Prediction," in *Conference: 4th International Conference on Computer Vision and Image ProcessingAt: MNIT Jaipur*, 2019, pp. 1–12. [Online]. Available: https://www.researchgate.net/publication/336530068

[14] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," International Journal of Forecasting, vol. 38, no. 4, pp. 1283–1318, Oct. 2022, doi: 10.1016/j.ijforecast.2019.06.004.

[15] C. Auppakorn and N. Phumchusri, "Daily Sales Forecasting for Variable-Priced Items in Retail Business," Proceedings of the 4th International Conference on Management Science and Industrial Engineering, Apr. 2022, doi: 10.1145/3535782.3535794.

[16] A. Keyaben Patel, N. Kumar, and S. Choudhari, "BigMart Sale Prediction using Machine Learning," *International Journal of Innovative Science and Research Technology* , vol. 6, no. 9, 2021, [Online]. Available: https://www.xajzkjdx.cn/gallery/423-april2020.pdf

[17] J. L. P. Ignatius, S. Selvakumar, S. JSN, and S. Govindarajan, "Data Analytics and Reporting API – A Reliable Tool for Data Visualization and Predictive Analysis," Information Technology and Control, vol. 51, no. 1, pp. 59–77, Mar. 2022, doi: 10.5755/j01.itc.51.1.29467.

[18] Bhavana T and Lakshmi K, "Machine Learning Algorithm for Predicting Big-Mart Sales," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 6, pp. 3457–3461, 2022, [Online]. Available: www.irjmets.com

[19] R. Dwivedi, "Sales Forecasting in Big Mart," Apr. 2020.

[20] R. F. Ali, A. Muneer, A. Almaghthawi, A. Alghamdi, S. M. Fati, and E. A. Abdullah Ghaleb, "BMSP-ML: big mart sales prediction using different machine learning techniques," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 12, no. 2, p. 874, Jun. 2023, doi:10.11591/ijai.v12.i2.pp874-883.

[21] Dr. G. S. Nana et al., "Machine Learning Approach for Big-Mart Sales Prediction Framework," International Journal of Innovative Technology and Exploring Engineering, vol. 11, no. 6, pp. 69–75, May 2022, doi: 10.35940/ijitee.f9916.0511622.

[22] A. Kothekar, M. Bodhale, P. Satapure, and R. Sarode, "Big Mart Sales Analysis Using Machine Learning," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT*, vol. 2, no. 5, 2022, doi: 10.48175/IJARSCT-4084.

[23] Y. F. Akande, J. Idowu, A. Misra, S. Misra, O. N. Akande, and R. Ahuja, "Application of XGBoost Algorithm for Sales Forecasting Using Walmart Dataset," Advances in Electrical and Computer Technologies, pp. 147–159, 2022, doi: 10.1007/978-981-19-1111-8_13.

[24] A. Arista, "Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19," Sinkron, vol. 7, no. 1, pp. 59–65, Jan. 2022, doi:10.33395/sinkron.v7i1.11243.

[25] R. Rawat and R. Yadav, "Big Data: Big Data Analysis, Issues and Challenges and Technologies," IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, p. 012014, Jan. 2021, doi:10.1088/1757-899x/1022/1/012014.

[26] P. Chhikara, N. Jain, R. Tekchandani, and N. Kumar, "Data dimensionality reduction techniques for Industry 4.0: Research results, challenges, and future research directions," Software: Practice and Experience, vol. 52, no. 3, pp. 658–688, Aug. 2020, doi:10.1002/spe.2876.

[27] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques," Measurement: Sensors, vol. 25, p. 100605, Feb. 2023, doi: 10.1016/j.measen.2022.100605.

[28] M. Mądziel and T. Campisi, "Energy Consumption of Electric Vehicles: Analysis of Selected Parameters Based on Created Database," Energies, vol. 16, no. 3, p. 1437, Feb. 2023, doi:10.3390/en16031437.

[29] Y. A. Alsultanny, "Big Data Visualization by MapReduce for Discovering the Relationship Between Pollutant Gases," Journal Port Science Research, vol. 4, no. 2, pp. 56–63, Nov. 2021, doi:10.36371/port.2021.2.3.

[30] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning,"

Energy, vol. 224, p. 120109, Jun. 2021, doi:10.1016/j.energy.2021.120109.
[31] D. Risqiwati, A. D. Wibawa, E. S. Pane, W. R. Islamiyah, A. E. Tyas, and M. H. Purnomo, "Feature Selection for EEG-Based Fatigue Analysis Using Pearson Correlation," in *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2020, pp. 164–169.
[32] T. Fu, X. Tang, Z. Cai, Y. Zuo, Y. Tang, and X. Zhao, "Correlation research of phase angle variation and coating performance by means of Pearson's correlation coefficient," Progress in Organic Coatings, vol. 139, p. 105459, Feb. 2020, doi: 10.1016/j.porgcoat.2019.105459.
[33] H. Pan, X. You, S. Liu, and D. Zhang, "Pearson correlation coefficient-based pheromone refactoring mechanism for multi-colony ant colony optimization," Applied Intelligence, vol. 51, no. 2, pp. 752–774, Aug. 2020, doi: 10.1007/s10489-020-01841-x.
[34] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," Social Science Research, vol. 110, p. 102817, Feb. 2023, doi: 10.1016/j.ssresearch.2022.102817.
[35] Pierce, Rod, 2024, 'Correlation', Math Is Fun, Accessed 19 Feb 2024. Available at: http://www.mathsisfun.com/data/correlation.html.
[35] A. Arista, "Visualization & correlation of big mart sales data," Portfolio of DSBIZ Certification. Accessed: Apr. 27, 2023. Available: https://bisa.ai/portofolio/detail/NjA4.