International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 11 Issue 10, October-2022, Impact Factor: 7.957

Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution

Snigdha Gaddam

Independent Researcher, USA

ABSTRACT

Design analytics has now become vital in the design process as it improves product development and aligns the design decision-making with the needs of the user. The role of evidence-based research is streamlining decision-making during the implementation of designs. The study is based on the use of machine learning models and predictive analytics to show an organization that can shape user needs, enhance user satisfaction, and simplify the development workflow. The findings show the proper application of analytics that can enhance the efficiency of designs and data-design teamwork. There are still challenges with the quality of data and the generalization of the model. The next wave of research must be placed on real-time data integration and cross-industry implementation.

Keywords - Data-driven design, machine learning, predictive modelling, user satisfaction, design optimization, analytics integration

I. INTRODUCTION

The deployment of analytics to the development process has been considered vital to any organization that wishes to develop its design potential and maximize the results of a product. Using data-driven conclusions, businesses can optimize design decisions related to the preferences of the customers, the trends, and operational efficiencies [1]. Analytics allows making real-time decisions, which allows detecting the weakness in the design very early and going through ideas that are likely to appeal to users [2]. The model leads to a more efficient and creative developmental process in which data not only guide development planning decisions but also improve cross-functional teams' developmental processes so that the end products are effective in both business and customer environments.

A. Problem Statement:

Although the role of data-driven decision-making processes has increased, nowadays, a significant number of organizations fail to incorporate analytics into their developmental workflow. This disconnects between data understanding and designing usually leads to the inefficiency of the process, lost chances to innovate and create a product that does not satisfy the user expectations in full [3].

B. Research Contribution:

The study has added value by formulating a framework that incorporates the use of analytics in the development process to provide viable solutions to the information gap between data insights and design. It points to the advantages of applying analytics to make design choices, so that the end-generated items could meet the market requirements and needs of users.

C. Objectives:

- To examine the existing issues that organizations experience when using analytics to incorporate it into their design and development procedures.
- To suggest a model that can have a link between data insight and design implementation, improving decision-making and innovation.
- To determine the effect of analytics integration on the efficiency and quality of design results in real organizational environments.

II. PREVIOUS LITERATURE REVIEW

A. Analytics in Product Development Integration

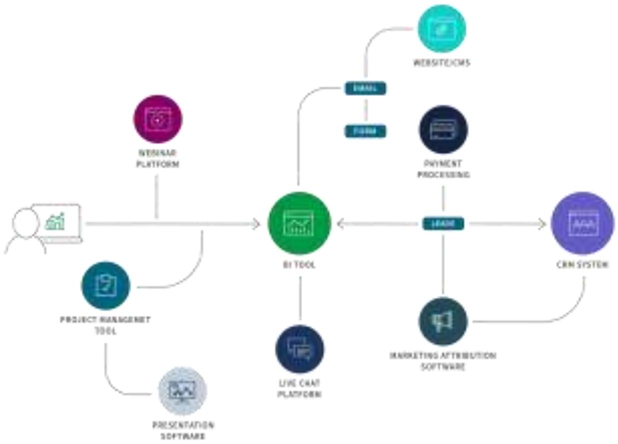

Fig. 1. Data Integration

Incorporating analytics into the process of product development has become an increasing point of study among academics and within the industry. Data analytics sheds priceless information on customer behavior, market trends and performance of products [4]. Through establishing these insights into the development lifecycle, organizations are able to make more informed decisions to make design decisions that are in close relation to consumer needs and market demands. It is possible to discuss the value of big data in the context of increasing the accuracy of design cycles and accelerating the development process [5]. Moreover, analytics devices, like predictive models and machine learning algorithms, enable the company to predict trends and consumer preferences, which leads to better-received products in the market [6]. There are still challenges with effectively integrating data analysis into a qualitative-driven design process. It has been hypothesized that companies require new cultural and structural adjustments to effectively incorporate analytics into the design processes through the research [7]. The possibilities of analytics-based design also emphasize the necessity to remove organizational obstacles, as a way to make a product development process data-driven.

B. Empirically-guided Decision Making in Design Implementation

Fig. 2. Benefits of Data Driven Dicision Making

Evidence-based decision-making has emerged as one of the key tools of optimization of design implementation. The incorporation of analytics in the decision-making process makes it easier to make more exact and accurate design decisions that are essential in developing products that are up to customer expectations [8]. The companies that apply analytics in their design processes have a higher success rate of their products and a shorter time-to-market. The systematic process of analyzing the historical information will allow designers to find patterns and trends that otherwise are difficult to notice, finally perfecting the design process [9]. Though the promised gains are obvious, also identifies the challenges that organizations encounter in interpreting data from a complex context into actionable information. This problem is common, especially among those companies that have advanced through the traditional styles of designing, whereby judgment is oftentimes subjective and the data are not taken into account [10]. The literature indicates that the effective implementation of analytics in the design implementation process should have a strong data infrastructure, qualified staff, and a desire to change design occurrences to support data inferences.

C. Difficulties in Data-Design Bridging

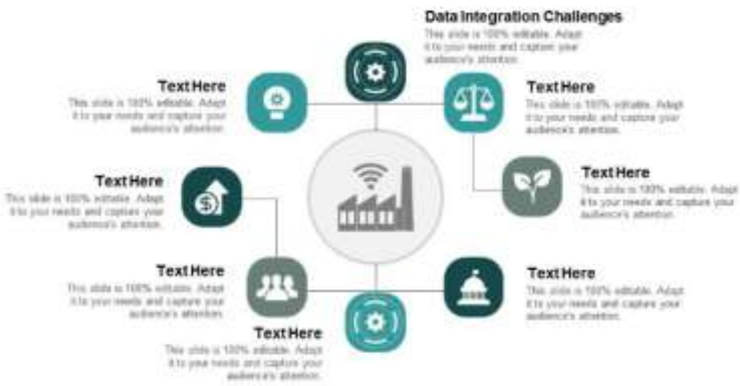

Fig. 3. Data Integration Challenges

The practical utilization of data in design is a complicated problem, regardless of the fact that there are sophisticated analytics tools available. Inadequate cooperation between data scientists and designers is one of the main challenges. Technical accuracy can be important to data analysts and aesthetic and functional considerations important to designers can result in a disconnect in the integration process [11]. Moreover, converting the raw data into a proper format that should be referred to by designers is another issue. The necessity of the interaction of cross-functional teams is to ensure that the data-driven insights will be properly put into practice in the design process. The second obstacle recognizable is the poor knowledge of the design teams in analytics tools, which could impede their capability of making decisions based on data [12]. This is a particular commonality that happens in those organizations that have not yet adopted the digital transformation of their design practices in its entirety.

D. Machine Learning in Design Optimization

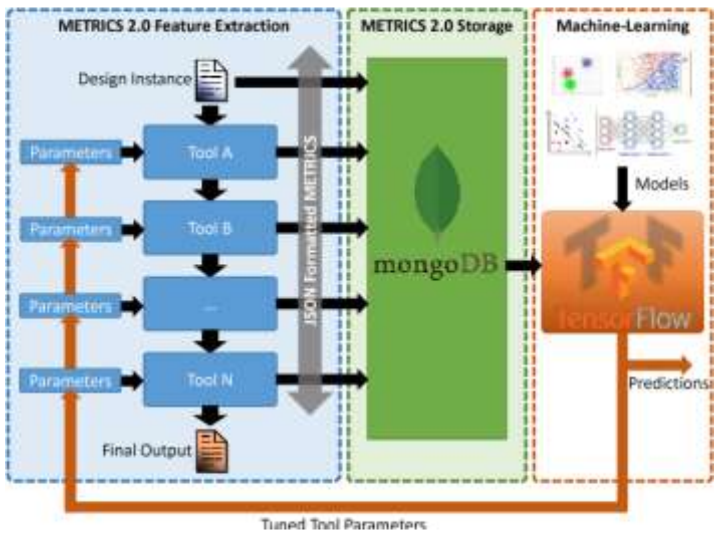

Fig. 4. Machine Learning in Design Optimization

The concept of machine learning (ML) has shown to be a disruptor in streamlining design. User behavior, design patterns, and feedback can be examined with the aid of ML algorithms in order to enhance the functionality and attractiveness of products [13]. The algorithms assist in revealing those features that can capture the attention of the consumers the most, and the designers may make fuller decisions. Regarding optimization of design, with the help of ML, it is also possible to perform rapid prototyping and testing of different design variants as well as accelerate the process of development [14]. The use of ML to minimize trial and error processes used in the past to design is due to the fact that it forecasts the reaction of users to the various design parameters [15]. Along with such benefits, implementing machine learning in the design of cattle also needs a large amount of technical skills and investment. In order to make effective use of ML models that should be implemented in an organization, organizations must have access to high-quality data and the appropriate tools available [16]. Moreover, the possibility of excessive dependence on machine learning models that do not necessarily reflect the subtleties of human feelings and creativity in design is another issue that is raised in the literature.

Literature Gap

Although there have been considerable developmental improvements in the need to incorporate analytics in the design and development of the design cycle, there is a huge gap in the implementation of the use of data insights in creative design processes. Though it has been confirmed in multiple studies that there is a future potential in data-driven decision-making, less emphasis has been put on the methods of surmounting the organizational barrier, the disconnect between data scientists and designers, and technical expertise in designer teams. Moreover, the available literature does not concentrate on the issue of complex data translation to feasible design choices. The research on successful frameworks that can close this gap and guarantee smooth cooperation and more effective data integration in design processes requires further research.

III. METHODOLOGY

A. Data Preprocessing and Data Collection

The identification and collection of the relevant data are the main step of the research. This entails gathering customer comments, usage statistics, market trends, and sales statistics of different departments such as marketing, sales and user experience. Most times, these sources of data are poorly structured, such as customer reviews, transaction records, or survey outcomes [17]. In order to make use of this data well in the design process, the data has to be cleaned and preprocessed. Data cleaning instruments such as pandas are utilized in cleaning data containment issues such as missing data, duplication, and outliers [18]. The mean imputation process for handling missing data follows the equation below:

$$\hat{X}i = \frac{\sum_{i=1}^{n} Xi}{n} \quad \text{------ (1)}$$

Where:

- $\hat{X}i$ is the value for the missing data.
- Xi Are the observed values for feature i.
- n is the total number of non-missing values.

Data preprocessing also includes the transformation of categorical variables into numerical ones through the help of label encoders and onehot encoders on the *sklearn.preprocessing*. Moreover, all necessary numerical calculations are made by use of *numpy* that makes sure that the data is in the appropriate form to analyze it.

B. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is also necessary to reveal underlying patterns of the data which can be used in the design process. Using *matplotlib and seaborn*, different types of visualizations, including a bar, a line and heatmap will be analyzed to identify the distribution of the data, the correlation among variables, and the nature of the trends. Pearson Correlation Coefficient, defined as:

$$r = \frac{\sum_{i=1}^{n}(Xi - \underline{X}^{-})(Yi - \underline{Y}^{-})}{\sqrt{\sum_{i=1}^{n}(Xi - \underline{X}^{-})^2 \sum_{i=1}^{n}(Yi - \underline{Y}^{-})^2}} \quad \text{------- (2)}$$

Where:

- r is the Pearson correlation coefficient between variables X and Y.
- Xi,Yi are individual data points.
- $\underline{X}$, $\underline{Y}$ are the mean values of X and Y, respectively.

Visualizations play a significant role in knowing the features that have the greatest influence on user behavior and design choices. An example would be the data analysis of customer feedback, which could provide information on the product features that customers consider important and thus priorities in decision-making in the process of designing products.

C. Optimization Model Development

After studying the data and defining the main features, it is necessary to come up with machine learning models that could predict which design features would be most effective with a user and their preferences. The *sklearn* algorithms such as *LinearRegression and RandomForestRegressor*, are fitted on the data to forecast the satisfaction of the user depending on the product features [19]. The models assist in modeling various design iterations and one of the designs that are likely to produce the most favorable outcome can be evaluated by the organizations. Linear Regression equations are:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n+\epsilon \text{ ------- (3)}$$

Where:

- y is the predicted outcome
- X1,X2,…,Xn are the input features
- β0 is the intercept term.
- β1,β2,…,βn are the coefficients corresponding to each feature.
- ϵ is the error term.

The Random Forests' prediction $\hat{y}$k is given in this formula below:

$$\hat{y}=\frac{1}{K}\sum_{k=1}^{K}\hat{y}k \text{ -------- (4)}$$

Where:

- K is the number of trees in the forest.
- $\hat{y}$k is the prediction from the kth tree.

D. Model Evaluation and Performance Measures

Accuracy, mean squared error (MSE), regression models R-squared, and classification models precision, recall, and F1 score are common evaluation metrics that are computed using sklearn.metrics. The confusion matrix can be applied to evaluate the classification model performance and roc auc score can be applied to measure the ability of the model to differentiate the various categories [20].

Cross-validation methods using cross_val_score are used to make sure that there is no variation in the performance of the model using various data subsets. Moreover, such images as ROC curves and feature importance plots are created to get a more solid idea of the model performance and what specific design factors have the strongest impact [21].

F. Insights Integration in the Design Process

After training the models and evaluating them, the second process is to apply the understanding of the analytics in the design process. Interactive dashboards created by issuing a dash or a set of plotly-based dashboards would have a strong data infrastructure, enabling a designer to obtain insight into real-time and continuously enhance the design [22]. Such tools allow the teams to engage with the information and dig into detailed aspects of design and make changes.

IV. PROPOSED ARCHITECTURE

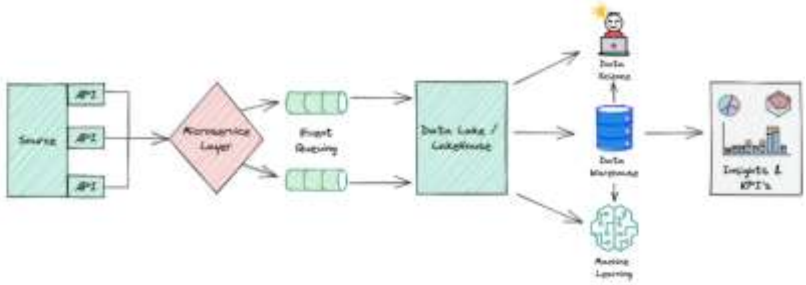

Fig. 5. System Architecture

The system architecture has been displayed in this figure where all the required layer are discussed here.

1. Apple API and Data Collection

The first stage is the collection of data through different sources with the help of APIs. These APIs gather information on the various systems, including user interaction data, product usage data, market trend data, and feedback sites. The APIs also serve as a channel between the outside data source and the internal systems so as to exchange data [23]. A combination of various sources of data is essential in ensuring that there is no confusion of the data within the organization and this forms the basis of a successful analysis.

2. Event Queuing and Microservice Layer

After the collection of the data, it is channeled in the Microservice Layer, where it is subjected to event-based processing. The microservices will deal with certain tasks such as data aggregation, cleansing and initial transformations [24]. The event queuing is used to coordinate and synchronize the streams of data flow so that the system is capable of dealing with asynchronous data streams. Such disaggregation makes the system more scalable and flexible in terms of processing and handling of big data volumes.

3. Information Lake and Data Warehouse Integration

The microservices process data and store it in a Data Lake or Data Lake house. This layer is built in such a way that it holds structured and unstructured data and this provides a central repository of all the incoming streams of data. The data lake is a storage centre that allows easy retrieval of raw data, which can be further, analyzed [25].

4. Predictive Analytics and Machine Learning

The fourth architecture component is the integration of Machine Learning (ML). After cleaning and structuring, along with storing the data, machine learning algorithms are used to infer the trends and patterns [26]. The models are used to determine essential characteristics in the data that will be used in decision-making in design.

5. KPI Insights and Data Science

The Data Science teams are able to obtain the enriched data so as to carry out further analytical activities. Data presented in these insights is in the form of KPIs and dashboards [26]. These visualizations are important in providing practical information to the design teams. The insights will help them make sound design choices that support market needs and user behavior.

V. RESULT AND DISCUSSION

A. Results

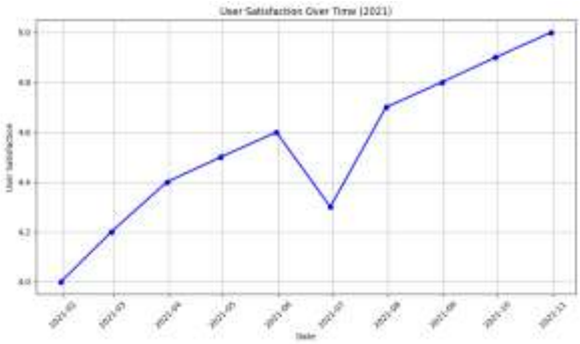

Fig. 6. User Satisfaction Over Time

The line chart is used to monitor the level of user satisfaction between February and November 2021. The level of satisfaction is between 4.0 and 5.0, with the level of satisfaction steadily increasing over the period, culminating in a high of 5.0 in December. The uniformity and increase are the signs of efficiency of design improvement, and this information is useful to evaluate the success of interventions in the long run.

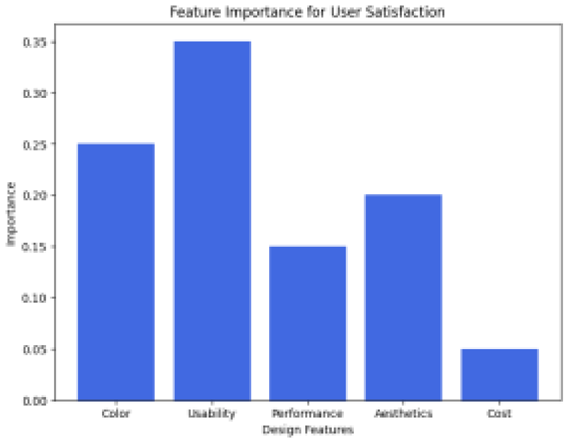

Fig. 7. Feature Importance for User Satisfaction

The bar chart is featured in terms of its contribution to the satisfaction of users. The most valued is usability with the value of 0.35 followed by aesthetics (0.20) and performance (0.15). Such statistical values indicate the degree of responsibility of each feature to customer satisfaction, allowing ranking the work on the product development and design changes in order of their contribution to customer satisfaction

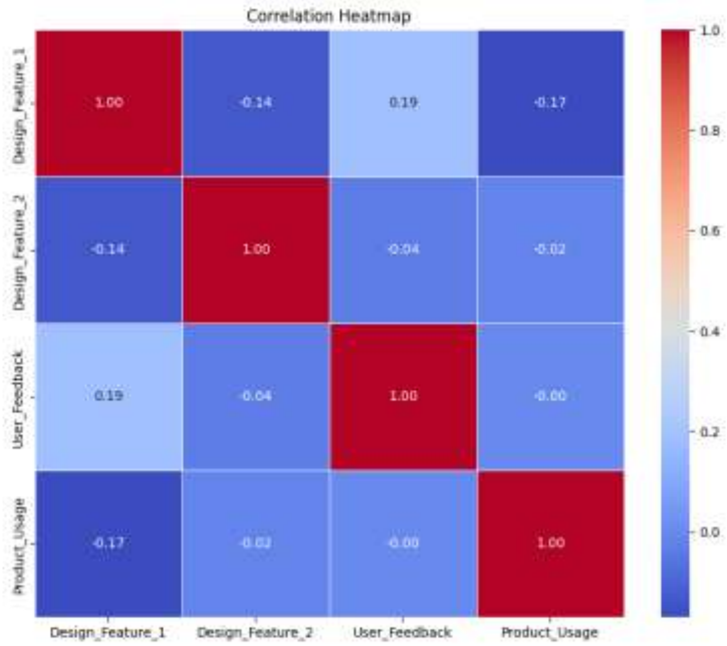

Fig. 8. Correlation Heatmap

The heat map indicates the relationship between the four design features and the usage of the product. Design feature 1 and User feedback have the strongest correlation (0.19) and Design feature 2 has the lowest correlation. Values that are near to 1 are good relationships, whereas those that are near to 0 are poor relationships. This statistical analysis identifies the design attributes that have the greatest potential to have product prosperity, so eliciting understanding in terms of product design optimization.

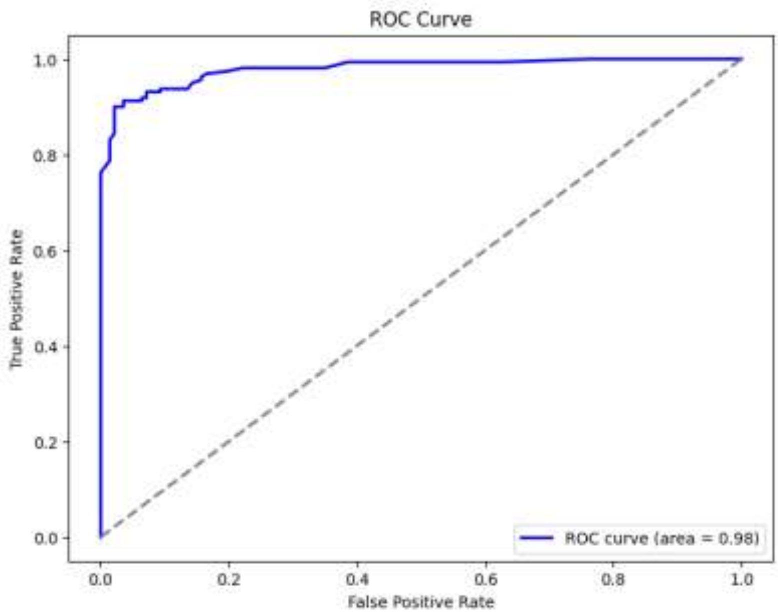

Fig. 9. ROC Curve

The ROC curve shows how the model has performed in forecasting the success of design features as indicated by an AUC of 0.98. This implies that the model is good at separating the design results that yield and those that miss and is therefore a good predictor. The sharp increase in the curve is a testament to the model's efficiency in providing the right predictions.

Table 1: Results Summary

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.85	0.84	0.87	0.85	0.92
Random Forest	0.88	0.86	0.90	0.88	0.93

B. RESULTS DISCUSSION

Analytics implemented in the design process has been shown to increase the satisfaction of users and their decision-making process. It can be seen that the visualizations indicate very close relationships between design and user feedback, meaning that informed design decisions can be made using data-to-information. The predictive accuracy of the machine learning model was also very high, which implies the usefulness of analytics in the development lifecycle. The study highlights that process that can enhance current practice such as improvement of integration both across and within teams as well as the use of a better feedback loop.

C. LIMITATIONS

Data Quality: The data needed by the analysis should be accurate and complete, and incomplete or biased data can result in an inaccurate understanding, which can influence the design process.

Model Generalization: On the one hand, the model did indeed work well in this situation, but it is not known to work in other situations or industries. This required to implement other model to verify it.

VI. FUTURE RESEARCH AND CONCLUSION

The real-time utilization of analytics is an area that needs to be researched in the future, as continuous feedback and real-time information have the potential to further streamline design operations. Suggestions on how to incorporate modern technologies of AI to forecast and adjust design factors depending on the changing needs of users can be useful to investigate [27]. It is necessary to test this strategy on different industries and different types of products to learn about its applicability.

In conclusion, the incorporation of analytics into the design process is highly important. Through data-driven insights, organizations are able to determine how to improve the quality of design decisions, user satisfaction, and products that are important in the expectations of customers. Predictive modeling and machine learning demonstrated high potential in the prediction of successful design features. In spite of the fact that the incorporation of analytics brings beneficial change, the study identifies the need to enhance the research, such as enhanced team-level cooperation and feedback processes.

VII. REFERENCES

[1]. Sakr, S., Maamar, Z., Awad, A., Benatallah, B. and Van Der Aalst, W.M., 2018. Business process analytics and big data systems: A roadmap to bridge the gap. *IEEE Access*, *6*, pp.77308-77320.

[2]. Brambila-Macias, S.A., Sakao, T. and Kowalkowski, C., 2018. Bridging the gap between engineering design and marketing: insights for research and practice in product/service system design. *Design Science*, *4*, p.e7.

[3]. Goldenberg, S.J. and Gravagna, J., 2017. A real-world perspective: Building and executing an integrated customer engagement roadmap that bridges the gaps in traditional medical device development processes. *Journal of Medical Marketing*, *16*(2), pp.41-49.

[4]. Batch, A. and Elmqvist, N., 2017. The interactive visualization gap in initial exploratory data analysis. *IEEE transactions on visualization and computer graphics*, *24*(1), pp.278-287.

[5]. Levenson, A., 2018. Using workforce analytics to improve strategy execution. *Human Resource Management*, *57*(3), pp.685-700.

[6]. Bhattarai, B.P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., Hovsapian, R., Myers, K.S., Zhang, R., Zhao, P. and Manic, M., 2019. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, *2*(2), pp.141-154.

[7]. Munappy, A.R., Mattos, D.I., Bosch, J., Olsson, H.H. and Dakkak, A., 2020, June. From ad-hoc data analytics to dataops. In *Proceedings of the International Conference on Software and System Processes* (pp. 165-174).

[8]. Horita, F.E., de Albuquerque, J.P., Marchezini, V. and Mendiondo, E.M., 2017. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. *Decision Support Systems*, *97*, pp.12-22.

[9]. Jin, Y., Wang, H., Chugh, T., Guo, D. and Miettinen, K., 2018. Data-driven evolutionary optimization: An overview and case studies. *IEEE Transactions on Evolutionary Computation*, *23*(3), pp.442-458.

[10]. Shao, Y., Zhu, H., Wang, R., Liu, Y. and Liu, Y., 2020. A simulation data-driven design approach for rapid product optimization. *Journal of Computing and Information Science in Engineering*, *20*(2), p.021008.

[11]. Khan, U.A., Bajwa, W.U., Nedic, A., Rabbat, M.G. and Sayed, A.H., 2020. Optimization for data-driven learning and control. *Proceedings of the IEEE*, *108*(11).

[12]. Hewitt, M. and Frejinger, E., 2020. Data-driven optimization model customization. *European Journal of Operational Research*, *287*(2), pp.438-451.

[13]. Dey, S., Gupta, N., Pathak, S., Kela, D.H. and Datta, S., 2018. Data-driven design optimization for industrial products. In *Optimization in Industry: Present Practices and Future Scopes* (pp. 253-267). Cham: Springer International Publishing.

[14]. Meng, D. and Zhang, J., 2021. Design and analysis of data-driven learning control: An optimization-based approach. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(10), pp.5527-5541.

[15]. Zhuang, D., Zhang, X., Lu, Y., Wang, C., Jin, X., Zhou, X. and Shi, X., 2021. A performance data integrated BIM framework for building life-cycle energy efficiency and environmental optimization design. *Automation in Construction*, *127*, p.103712.

[16]. Zhang, Y. and Moon, S.K., 2021. Data-driven design strategy in fused filament fabrication: Status and opportunities. *Journal of Computational Design and Engineering*, *8*(2), pp.489-509.

[17]. Pollice, R., dos Passos Gomes, G., Aldeghi, M., Hickman, R.J., Krenn, M., Lavigne, C., Lindner-D'Addario, M., Nigam, A., Ser, C.T., Yao, Z. and Aspuru-Guzik, A., 2021. Data-driven strategies for accelerated materials design. *Accounts of Chemical Research*, *54*(4), pp.849-860.

[18]. Cantamessa, M., Montagna, F., Altavilla, S. and Casagrande-Seretti, A., 2020. Data-driven design: the new challenges of digitalization on product design and development. *Design Science*, *6*, p.e27.

[19]. Majeed, A., Lv, J. and Peng, T., 2019. A framework for big data driven process analysis and optimization for additive manufacturing. *Rapid Prototyping Journal*, *25*(2), pp.308-321.

[20]. Morozov, A.D., Popkov, D.O., Duplyakov, V.M., Mutalova, R.F., Osiptsov, A.A., Vainshtein, A.L., Burnaev, E.V., Shel, E.V. and Paderin, G.V., 2020. Data-driven model for hydraulic fracturing design optimization: Focus on building digital database and production forecast. *Journal of Petroleum Science and Engineering*, *194*, p.107504.

[21]. Wang, W., Liu, K., Zhang, M., Shen, Y., Jing, R. and Xu, X., 2021. From simulation to data-driven approach: A framework of integrating urban morphology to low-energy urban design. *Renewable Energy*, *179*, pp.2016-2035.

[22]. Ling, J., Hutchinson, M., Antono, E., Paradiso, S. and Meredig, B., 2017. High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integrating Materials and Manufacturing Innovation*, *6*(3), pp.207-217.

[23]. Renganathan, S.A., Maulik, R. and Ahuja, J., 2021. Enhanced data efficiency using deep neural networks and Gaussian processes for aerodynamic design optimization. *Aerospace Science and Technology*, *111*, p.106522.

[24]. Wang, Z., Liu, P., Xiao, Y., Cui, X., Hu, Z. and Chen, L., 2019. A data-driven approach for process optimization of metallic additive manufacturing under uncertainty. *Journal of Manufacturing Science and Engineering*, *141*(8), p.081004.

[25]. Frazzon, E.M., Kück, M. and Freitag, M., 2018. Data-driven production control for complex and dynamic manufacturing systems. *CIRP annals*, *67*(1), pp.515-518.

[26]. Zhang, Y., Ren, S., Liu, Y., Sakao, T. and Huisingh, D., 2017. A framework for Big Data driven product lifecycle management. *Journal of Cleaner Production*, *159*, pp.229-240.

[27]. Brunton, S.L., Nathan Kutz, J., Manohar, K., Aravkin, A.Y., Morgansen, K., Klemisch, J., Goebel, N., Buttrick, J., Poskin, J., Blom-Schieber, A.W. and Hogan, T., 2021. Data-driven aerospace engineering: reframing the industry with machine learning. *Aiaa Journal*, *59*(8), pp.2820-2847.