Leveraging Machine Learning for Predictive Analytics in U.S. Aerospace Manufacturing: Techniques and Case Studies

Dr. Krzysztof Kowalski

Associate Professor of Computer Science, Warsaw University of Technology, Poland

1. Introduction to Machine Learning in Aerospace Manufacturing

Aerospace manufacturing provides a real-world, high-stakes environment to experiment with predictive analytics, and more broadly in leveraging machine learning. The need to reduce dependence on legacy algorithms and manual operations in order to integrate machine learning into predictive analytics drives the motivation of this paper. We present a comprehensive review of the applications of a variety of machine learning algorithms within the context of aerospace manufacturing. Through this, we summarize the body of existing work and provide a detailed account of other successful approaches to the application of deep learning within predictive analytics.

Given such an emphasis on defense and commercial concerns in regard to this manufacturing anomaly and its potential to bring harm, in this contributing essay, we will focus strictly upon the integration of machine learning in predictive analytics throughout U.S. aerospace manufacturing industries. We focus on this application of machine learning in predictive analytics as almost the entire domain of machine learning research is dedicated to the forecasting of outcomes based on input. We aim to understand what connects machine learning to predictive analytics by focusing specifically on its application in the aerospace manufacturing industry to propose this linkage. The size of the U.S. aerospace manufacturing sector and its role in developing the innovation and technology of tomorrow demands that any improvement within these operations have detailed analysis and data to illustrate it. This essay argues the relevancy of machine learning within the aerospace manufacturing domain and provides multiple case studies in a variety of roles and organizations throughout competition.

1.1. Overview of Predictive Analytics

Predictive analytics involves analyzing existing data to make predictions about future events or outcomes. Various statistical or machine learning techniques can be used for this purpose. Applications of predictive analytics in aerospace manufacturing typically include predicting drilling thrust, tool life (e.g., delamination factor, chipping factor), surface roughness, and shop floor measurements, among others. Aerospace manufacturing processes are, by nature, costly (manufacturing a small portion of an airplane might cost more than a billion US dollars). Thus, reducing waste concerning raw materials, time, and energy results in significant savings, and developing predictive systems can contribute to this end. In a broader context, introducing machine learning systems in aerospace implies better dynamic control. These predictive systems also deliver competitive advantages, i.e., the improved acceptance states of workpieces in machining, since only those that have been fabricated close enough to nominal geometry and require less rework are going to move on and typically leave less time between the final assembly stages.

Most of the techniques for predictive analytics rely on a machine-learning approach, and the review is preceded by explaining their original machine-learning paradigms. With the advent of Big Data tools and technologies, modern predictive analytics has transformed into data-driven learning and can therefore be characterized using a machine-learning context, rather than the classic statistical methods. In aerospace manufacturing, machine learning has found various niche applications. This includes, but is not limited to, using electrochemical impedance spectroscopy to detect damage to CFRP, material processing (one-shot manufacturing), using machine learning for controlling and improving the quality of additively manufactured parts, and using neural networks to classify "good" versus "bad" carbon fiber textiles.

2. Fundamental Machine Learning Techniques

In machine learning, we have the following fundamental learning techniques that comprise the core building blocks used to create amazing, exciting, and more advanced learning methodologies that are particularly relevant for aerospace manufacturing: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is a machine learning technique that is generally used to make predictions of unknown future outputs or events. The training dataset available has features (or input vectors) and target values (or ground truth). Using these datasets, we can create a model that when given an unseen data/feature, will make a prediction of outcomes. The main application is classification or regression, which is for estimating continuous value outputs, e.g., demand forecasting and percentage scrap prediction. Technique: artificial neural networks (ANN), support vector machine (SVM), decision trees (DT), bagging, boosting, random forest, gradient boosting.

In contrast, unsupervised learning techniques have no target/output values on what we can train our machine learning algorithm. The technique chiefly deals with clustering the products or potential customers into different categories based on their attributes, e.g., k-means clustering, hierarchical clustering. These techniques are further used for the reduction of redundancy in the available dataset and for reducing the total number of data attributes. For instance, Principle Component Analysis (PCA) is one of the methodologies used to reduce the number of data attributes based on the condition that the objective of the attribute reduction is data structure preserving as much as possible.

Finally, reinforcement learning adopts strategies based on elimination factors using a system that changes in response to the performance of the agent while interacting in the environment to obtain the best possible reward. It typically involves multiple iterations for arrival at the best solution through the process of continuous learning, e.g., usage in simulators.

2.1. Supervised Learning

Supervised learning is a foundational machine learning technique and an important part of predictive analytics. In addition to simply categorizing things, supervised learning can be used for predicting a discrete or continuous target based on multiple input variables. In the case of predicting multiple continuous targets, supervised learning can be extended to become a multivariate model.

Supervised learning is best understood by looking at classifier and regression applications, which are described further in the support vector machines and linear regression sections, as well as applications in the predictive analytics case studies of Sections 2.3 and 2.5. In the context of supervisory knowledge, the object suggestion cloud in Fig. 4a and the part-avoidance suggestion cloud in Fig. 5a form the notion of a feature cloud and a target cloud, respectively. A feature-target pair would represent a single dot in the resulting 3-D plot for part-avoidance prediction. For the fraction of points in these clouds, the target is known as the supervisory label. Beyond the question of effort, there are several theoretical reasons why it might be either impossible or infeasible to retrieve such information. In this section, the basic tenets of classifier and regression applications in machine learning are discussed. Techniques for the identification of error within these applications are also introduced.

2.2. Unsupervised Learning

Unsupervised learning encompasses a plethora of machine learning techniques that seek to distill knowledge and identify patterns in data by extracting meaningful features embedded in raw data, without supervision or pre-defined labels. In contrast, supervised learning relies on labeled data where instances of the training data have associated outputs, or labels, making it easier for the model to adjust. Unsupervised learning techniques are particularly attractive for the aerospace manufacturing domain due to the multi-modal nature of the data – data can be generated at multiple subcomponents, from different stages of production, or by using different pulverization methods of metallic systems, to name a few modalities. It is also difficult to disentangle all relevant causes for degradation (single, or combination of many) and to benchmark spent nuclear fuel used for calibration campaigns to alternatives manufactured at different facilities.

Innovative data-processing tools falling into this category can therefore be of interest for rapidly and exhaustively understanding chemical and morphological heterogeneity. Principal component analysis, or PCA, is a frequently used unsupervised tool. It is based on linear algebra and is used to reduce the dimensionality of the data into a few principal components that explain the majority of the data's variance. The transformed data, i.e., after projecting its dimensionality is reduced, can be used for further analysis, such as clustering or for use of a predictive model. Of marked interest for performance enhancement is clustering data to identify trends in the data, or to identify distinct groups of samples. A variety of clustering algorithms exists, each utilizing different mathematical approaches to cluster samples into two or more clusters based on some mathematical minimization or maximization function (e.g. minimizing intra-cluster variance, or maximizing inter-cluster distance). Popular clustering algorithms include k-means or the agglomerative hierarchical clustering or DBSCAN.

2.3. Reinforcement Learning

Reinforcement Learning: Reinforcement learning (RL) is another fundamental machine learning technique. RL approaches have been mainly popular in gaming and robotics but are now being increasingly used in problems where the objective is to optimize a trade-off between exploration and exploitation. In a typical RL scenario, the machine agent interacts with the environment, samples the performance of a decision in a given scenario, learns from the performance, and then takes the next decision. RL can be a potential area of exploration in an aerospace manufacturing setup where the quality of the product depends on a sequence of decisions and state. The sample applications in the aerospace manufacturing setup can be job sequencing, tool-path optimization, autonomous inspection and assembly lines, and energy control or predictive maintenance and logistics.

Advanced Techniques for Predictive Analytics in Aerospace Manufacturing: In the backdrop of these fundamental techniques, one can explore more advanced techniques such as advancements in reinforcement learning, adversarial networks, hierarchical deep learning, and deep similarity learning for applications in aerospace manufacturing. We cover these advanced techniques in the subsequent scenario-based explorations sections. In this section, we have presented a brief background of various techniques that can be useful for predictive analytics. The techniques range from rule-based and optimization to different machine learning and deep learning-based techniques, which can be used in a discrete as well as a continuous analytical setup. We will now discuss a few relevant use cases in aerospace manufacturing illustrating the utility of advanced predictive analytics methodologies.

3. Advanced Machine Learning Techniques

The field of machine learning has experienced a significant boom in the last five years. Consequently, more advanced modeling techniques have also been established. While the traditional, commonly used methods such as ARIMA and Random Forest have become the topic of numerous research publications over recent decades, discussing advanced topics while simultaneously accounting for published content on the matter is challenging. To define contents in such a manner that both remain an integral next step for implementation and avoid repeating existing efforts, studies targeting the manufacturing industry are becoming more and more popular. A growing trend in employing deep learning models and an ensemble approach has arisen in forecast activities in the aforementioned sector. With respect to other manufacturing industries, limited research has been dedicated to the use of these methods in the context of aerospace.

Deep learning has been successfully deployed in aerospace engineering (i.e. for predicting future rewards for control problems). Furthermore, multiple ANN layers have been used to develop forecasting models in complex manufacturing, enabling the capture of diverse representation features. Ensemble learning, on the other hand, has gained more attention in recent years in the literature concerning the aerospace industry to tackle quality monitoring, predict demand of aviation, surface quality of products and lifetime remaining of machine tools. In this paper, the advanced forecasting techniques will be discussed in detail and proposed in novel use-cases catering to aerospace.

3.1. Deep Learning

Deep learning (DL) refers to a group of machine learning algorithms that perform a variety of tasks, including supervised (e.g., classification, regression, and ranking), unsupervised (e.g., clustering and dimension reduction), or hybrid learning (e.g., reinforcement learning, sequence generation). Deep learning is appealing to engineers and scientists because of its proven performance and generalizability across various sectors and applications. Importantly, deep learning advances support predictive analytics in innovative ways. Rather than applying handcrafted rules or features to the data, the integration of deep learning approaches facilitates automatic extraction of meaningful, and often latent, representations from data. The level of performance gains depends on the availability of data as well as the related problem and associated application. The final resulting repertoire has the potential to revolutionize the predictive analytics and decision-making process, often identifying factors and variables that carry important content that add new knowledge and insights that may be leveraged for improving aerospace manufacturing performance. Some applications within aerospace manufacturing have been reported, including smart material selection in hybrid composites processing and pinpointing the source of variations to improve manufacturing of next-generation composite-intensive aircraft.

Depending upon availability and the degree of uncertainty that an organization may incur, a wide variety of deep learning architectures is accessible to practitioners. Figure 6 provides a graphical representation that characterizes some common variants within the different categories of deep learning. More detailed insights into various deep learning approaches, such as the multilayer perceptron, convolutional neural networks (CNN), recurrent neural networks (RNN) – particularly long short-term memory (LSTM) and gated recurrent units (GRU) – attention mechanisms and transformers may be accessed in the Appendix. To gain traction in the smart manufacturing domain, access a recent comprehensive review on the practical aspects of deep learning.

3.2. Ensemble Learning

Ensemble learning is an advanced machine learning technique, widely used and well-suited for aerospace manufacturing. Its primary objective is to leverage individual prediction "experts" to achieve a superior aggregation outcome. Developing a multitude of learning "experts" can result in accurate predictions with minimal bias/variance, which are serious challenges associated with real-world modeling of manufacturing data and processes. Ensemble learning promises the combination of weak/strong predictions generated by i) multiple machine learning algorithms (heterogeneous ensembles) to improve robustness against modeling errors, ii) different training subsets or t-time glimpses of the data (homogeneous ensembles) to incorporate collective learning and tracking temporal dynamics and drifting behaviors of solutions over time, and iii) varying model representations and predictions generated by the same algorithm with different hyperparameter values (randomized ensembles) to exploit bias-variance trade-offs across multiple base learners. The most common and computationally efficient type of ensemble, both homogeneous and heterogeneous, is a bagged ensemble of trees.

The choice of ensemble learning over a single learning strategy should involve: a) identifying a learning task or environment that requires different experts to interpret inputs differently and/or learning long-term dependencies by tracking and exploiting the outcomes of different patterns, b) constructing a large number of learning functions/predictions, each focused on either disjoint subsets of the input features, disjoint subsets of the training data, different time windows and feature views in multiple training rounds, or using different hyperparameter settings, c) combining predictions using some mechanism to incorporate collective learning of the learner experts that can also evaluate individually the reliability of other experts. These steps, despite presenting better model performance, can be computationally exhaustive.

4. Data Preprocessing in Aerospace Manufacturing

Data preprocessing is one of the most crucial steps in building and deploying machine learning models in a wide range of applications, including the realm of aerospace manufacturing. The concept of data preprocessing is best described through the machine learning pipeline. The first step of the pipeline consists of data retrieval. The second step of the pipeline is data preprocessing, in which the dataset is prepared for model training. Once the dataset is preprocessed, the third step of the pipeline is model training. The trained model is then validated (model evaluation as well as model selection) using a validation dataset. As part of data preprocessing, most datasets require some sort of feature engineering. The feature engineering process consists of data cleaning, normalization and standardization, or determining item specifications.

Feature engineering is a creative process, and generally is not covered in great detail in publications. Research articles often provide a high-level description of the feature engineering process for a specific problem, but they do not contain the specific steps it takes to adequately prepare a dataset for model prediction. It is the responsibility of the practitioner to decide which of the techniques highlighted herein will work best for a given set of data. Aerospace component reliability is a complex feature that consists of a number of single-valued properties. These properties are both materialistic and performance-based. Materials that exhibit strength in tension and fatigue resistance also have a premium cost over less capable materials. This balance forces aerospace design teams to consider many trade-offs before constructing a machine learning model. Data preprocessing plays a crucial role in preparing heterogeneous datasets for model induction. Each of the techniques highlighted is part of a standard methodology to create high-quality, reliable predictive analytics.

4.1. Feature Engineering

Feature engineering is of paramount importance in preparing data for machine learning (ML) applications. Data in, for instance, aerospace manufacturing often comprise time series with continuous, multi-modal, and nonlinear trajectories; thus, inherent unique domain knowledge is critical for feature selection, generation, and transformation. Low-level features may not be meaningful on their own for certain univariate prediction problems, and 'disguise multimodal data' must be aggregated from which high-level statistics may be more informative. Vector and visual time series are common in practice; the former can be transformed to tabular (structured) data to describe the distribution of values or characteristics of samples, such as shape (e.g., kurtosis), dispersion (e.g., standard deviation), symmetry (e.g., median), centre (e.g., mean), and density (e.g., range); the latter can be transformed by extracting symbols (with sliding-windowing) to a categorical sequence.

New features may also result from domain knowledge, heuristic methods, and principal and independent component analysis. High-order features can also be generated with temporal event data through, for instance, time-delayed embedding via windowing to capture hierarchies, patterns, and trends in sequences of data for better time and frequency representation. Therefore, new higher-level representations, transformations of raw data, and valuable ancillary information can be taken as strong predictors in forecasting problems, based on descriptive and prescriptive analyses that dose response, inform on the process state, and capture the detection time and required future sampling rate for events, and can be employed in condition-based and predictive maintenance and quality assurance, which in turn may result in improved decision-making with human-in-the-loop (HyML). Time-series data that describe machine degradation, health status, performance, and behavior in aerospace manufacturing applications are actually feature engineered; the focus is now on novel methods for automatic—though guided—feature engineering.

4.2. Data Cleaning

Data cleaning or preprocessing is the essential verve of smarter predictive models. Just as thwart and on occasion, machinery of any type will yield observable and measurable information about past and relentless states. When this information is harnessed and used to train machine learning algorithms, they are capable of elucidating relationships and bringing to light patterns not discernible to even the keenest technicians. Clean training data is not only empirically sound, it is also morally responsible. It ensures fair, accurate, and consistent outcomes by creating predict.

Fig. 2: Nonlinear model for the aerospace component repair station. 1. Current state according to various three indicants. 2. Start of the nonmeasurable crane crack. 3. Future through extrapolation of the two measured indicators. 4. Nonlinear model for the manned assembly process (e.g. OSIRIS-REx spacecraft) and best results. 5. Impact of the three indicators after hand-coding the two fuzzy memberships for precision = 75%, recall = 67%. 6. Significance of the same three indicators in OSIRIS-REx. Values are corrected to be a percentage of the normal USL of each indicator. 7. Correlation after comparing component candidates and learning

from the moon mining data with multiple crafts. 4.5 Data Integration and Knowledge Management Integration of aerospace datasets is noted to be fraught with challenges such as ring-fencing. To date, existing approaches to aerospace data integration do not always encompass the automation of unidentified problem area segmentation. The integration of manufacturing knowledge is critical when considering the immutable accuracy/precision in predictive models. Consequently, half a decade ago an additional approach was required to include the outlook of collaboratively-managed information discovery for aerospace manufacturing. Knowledge management was embraced to move away from research considered with singular techniques. The results ascertained predicted an aerospace automated hyperformance hybrid shifty system whose performance and onset of fault can be predicted.

4.3. Normalization and Standardization

In this aerospace manufacturing domain, normalization concerns transforming a feature to a new scale, such as [0,1] or [-1,1]. Feature scaling techniques can benefit the preprocessing stage of data, aiding in the reduction of the model's ordering-related biases. Adapting feature distributions can potentially hold the subsequent machine learning model to be simpler. Additionally, within the context of the proposed predictive maintenance field of exploration, normalization and standardization can support model convergence during the training part of the neural networks, such as stochastic gradient descent, and influence the timeliness of the algorithm's learning process. Therefore, the benefits of normalization and standardization for this study include the processing of our data for predictive analytics, especially considering that an array of data preprocessing methods assist in the improvement of model evaluation and selection.

Normalization involves bringing the feature to a similar scale, such as [0,1] or [-1,1]. The main goal is to standardize the range and maximum/minimum. Possible advantages: algorithms that need weights (e.g., Neural Networks, Stochastic Gradient Descent) converge faster, and the range of feature values is simplified. Standardization implies the adjustment of feature values to have a mean value of 0 and a standard deviation of 1. Possible advantages: Algorithms that rely on features' "distributions" (e.g., using the Euclidean distance) are impacted, and Boosting may be enhanced if an optimal value for a weak learner is sought. In the next section, based on the above discussions, first, the U.S. aerospace manufacturing industry is introduced. Then, a variety of predictive analytics studies are summarized in this field, with a particular emphasis on machine learning and deep learning algorithms.

5. Model Evaluation and Selection

Model Evaluation and Selection: The performance of a supervised classification model is evaluated based on its ability to identify physical defects in aircraft parts as "Acceptable" or "Rejectable." Poor performance may result in excessive system downtime or, worse, missed detection for faulty parts. As a result, the choice of light GBM as the final model is influenced by both practicality and the performance of the model. In the dataset, the class distribution is rather imbalanced. As a result, cross-validation is used to validate all our model assessments. Stratification assists in ensuring that the validation sets retain the same percentage of accept-reject class distributions as the original dataset.

Performance metrics: The ROC curve plots the TPR (True Positive Rate or Sensitivity) against the FPR (False Positive Rate) (1-specificity) for all distinct thresholds. The best model is one that has more true positives and fewer false positives. As a result, an acceptable model achieves a high TPR and a low FPR, with the threshold chosen during the modeling process. The AUC score, which calculates the area under the ROC curve, gives a singular value to evaluate the model. If the AUC score is large, it is assumed that the model is good at differentiating between the two classes. Including the AUC value for both training and validation datasets allows us to see not just the model's capacity for better fitting (training) but also its potential to perform in a fresh dataset (scan).

5.1. Cross-Validation

Cross-validation. Cross-validation is a resampling technique used in machine learning to assess the performance of a model. More than evaluating the model for overfitting, underfitting, bias, and variance, cross-validation can be used for estimating and obtaining the performance metrics such as Area Under Curve (AUC), accuracy, precision, recall, and F1 score. The evaluation metrics quantifying the relationship between true labels and model predictions are subsequently used to ascertain the generalizability of the model. For instance, to verify the ability of predicting time-to-failure in a simple logistic regression model, airlines' historical data may be used for validating the possibility of predicting upcoming close to failure times. Definitions for precision, recall, F1, AUC, and accuracy can be found in Dean et al. 2013.

Cross-validation techniques, together with a thorough analysis of error distributions in the derived outcome measures, are essential in deciding how to develop/tune follow-on algorithms in the predictive maintenance domains. For example, given a predictive-improvement-idea and existing pilots, one must combine blind-pilot tests, as if they were completely new data, in rigorous cross-validation holders. Defining those using precise domain knowledge and targeting imbalance in the failure distribution. This section starts with the cross-validation techniques and then further narrates some of the most important techniques, namely k-fold validation, time series split, and grid search. Bosch's evaluation method is also briefly discussed before possibly describing case studies out of the domain of aerospace manufacturing.

5.2. Performance Metrics

Performance metrics are a critical part of predictive analytics for aerospace manufacturing. In the manufacturing environment, potential performance metrics can include, but are not limited to, cost as a percent of sales or in absolute terms, delivery time, flexibility or time to change quickly, ability to manage the network from outside, time to market, and track record. A leader can use performance metrics to distinguish machine learning model selection. In this paper, we provide a comprehensive overview of the specific performance metrics within the sphere of aerospace predictive analytics.

Predictive analytics pertains to the extraction of information from data to be used in forecasting future and/or identifying optimal resource allocation. Predictive analytics utilizing machine learning involves the application of algorithms to large datasets to extract important relationships, evaluate these relationships for consistency, and, recursively, test model performance as new data become available, before deeming the model ready for deployment. Specific performance metrics used to validate a machine learning model are based on subjective inference. Prediction error, for example, can be measured in terms of percent; using Eq. (4), one can calculate the percentage of the two-year prediction error which is 6.0 percent. In the world of aerospace manufacturing, the following performance metrics may be considered:

1. Discrete part failure as a percentage of time. 2. Discrete part failure as a percentage of products and/or services. 3. Average interim between part failure, for each prediction horizon, e.g. 3 months, 6 months, and 12 months.

The following performance metrics can also be considered in aerospace manufacturing using machine learning:

1. The mean squared error (MSE), or in a predictive maintenance organization, need to consider the side of the predicted error e.g.: MSE of the number of years until next failure per discrete machine part. 2. Absolute percentage error (APE), or similarly mean percentage error (MPE). 3. Area under the curve. 4. R-squared, e.g., the correlation coefficient, over time the model took a hit of "r-squared" dropping to less than 0 at "d" of 48, when the national economy took a dramatic impact, and remaining less than 0 for one year, and at "d" 66, the startup and bankruptcy of new information companies reinitiated a correlation index to the national GDP, increasing "r-squared" to about 1. 5. Precision: guide manufacturers to decide which assets to focus on based on which equipment are working well and unpredictably. 6. Or recall: recall is the metric for filmmakers using predictive maintenance recruitment strategies.

The purpose of this paper is to provide background on the proposed predictive analytics applications and provide a comprehensive review of the prospective techniques. In this regard, we present a four (04) part paper. This is Part I, where we presented our use case. In Part II of the paper, we will test a variety of machine learning techniques in both application domains and provide a comprehensive discussion of results, comparing and contrasting the three machine learning algorithms and their computational complexity statistics. We will conclude by identifying the machine learning algorithm that, based on a variety of performance metrics, would be the most suitable for scale-up to a predictive analytics application. In Part III, we will explore clustering applications in both domains, e.g., grouping different types of firefighter fatalities, and identifying different severity groups for road crashes involving bicyclists. We will then compare and contrast clustering results of the provided police reports to application-specific sources. Finally, in Part IV, we will present conclusions and future research directions.

6. Case Studies in Aerospace Manufacturing

Estimates show that the aircraft maintenance industry in 2015 had revenues of about $60 billion, with 45% (about $27 billion) of the revenues of the maintenance sector derived from maintenance and service training and support. For airlines, the cost of aircraft sector maintenance significantly impacts overall operational costs as well as the workforce required to maintain their fleet of aircraft. As such, more data-informed approaches to economic gate-to-gate air travel are being explored by the Federal Aviation Administration (FAA) Joint Planning & Development Office as well as the National Aeronautics and Space Administration. One promising approach to understanding the entire logistics ecosystem that might eliminate delays and improve passenger travel is a thorough focus on the aviation maintenance industry.

A prime area of exploration is in the analysis of the aircraft component overhaul arena. The main business process for component MROs typically begins by procuring components that are known to be in demand in the aftermarket. Due to the explosive growth and increasing automation in the US aerospace industry, the focus of this section will be limited to predictive maintenance and predictive quality control practices that directly impact the component MRO process. In the following paragraphs, we describe some of the key predictive analytics efforts that are currently underway in the context of the US aircraft manufacturing industry.

6.1. Predictive Maintenance of Aircraft Components

Predictive maintenance is appropriate when a cost-effective paradigm shift in maintenance policy is desired. For instance, the maintenance industry has traditionally been driven by procedures that attempt to predict the potential for failure. Condition-based maintenance (CBM), where maintenance is carried out on the basis of the measured condition of the item or its environment, is such a method. This allows an organization to see how its critical assets are performing in real time. The underlying principle of CBM is root cause analysis: addressing the problem at the source results in fewer corrective problems. Thus, CBM provides information on any potential problems in order to speed up the maintenance process and reduce downtime. This can proactively extend the life of the asset and reduce the downtime for certain equipment. It is important to note that the application of CBM can be limited to mission-, safety-, or business-critical equipment or situations.

Aerospace manufacturing—a sector in the U.S. that generates profits from producing military aircraft and their components—can benefit heavily from the advances outlined above. In this section, we focus on a case study conducted in collaboration with scientists and engineers at Arnold Engineering and Development Complex (AEDC), a test facility of the United States Air Force (USAF). AEDC is responsible for testing the reliability and performance of fighter aircraft, jet engines, space propulsion systems, turbines, and other equipment utilized in aerospace manufacturing to ensure that they are dependable and function as expected. In this case study, we use maintenance data of critical components of aircraft tested at AEDC to investigate failure prediction. The goals are to automatically detect features that correlate strongly to an ongoing failure and are computable in real time, obtain a failure prediction before the test gets damaged, and utilize change-point analysis to validate the SWaT system's condition. The main contributions are to (a) provide an example of applying machine learning techniques in a real-world operations and maintenance context, and (b) demonstrate the value of failure prediction when properly translated into action can affect revenue streams. To the best of our knowledge, no one has previously presented failure operational data in the open literature in aerospace manufacturing.

6.2. Quality Control in Aerospace Manufacturing

Quality Control (QC) applications in U.S. aerospace manufacturing are an area of interest. As of 2016, more than 700 standard measurement and inspection procedures are published for the U.S. aerospace program known as United Launch Alliance's (ULA) Launch Vehicle programs. The incentive for investing in these QC applications in aerospace is high due to the expense of errors. Handmade quality processes may yield redundant checks leading to additional time spent waiting on quality assurance. As patent restrictions regarding the detail of many aerospace designs are still valid, not only the outputs of production lines but also operations and maintenance processes are confidential. Therefore, research on learning from data in aerospace manufacturing operations and maintenance are generally underrepresented compared to consumer-oriented systems or business processes. The complex aerospace components LPSC produces for NASA also require several tens of hours of labor, making the units produced in a single day relatively small. This paper seeks to provide a detailed investigation of machine learning for predictive analytics in a very specific and intensive human capital-oriented quality control setting. The research question intends to highlight which techniques are most commonly used, as well as highlight any strongholds and pitfalls for this specific audience.

Specifically, quality control for the Liquid Propulsion Systems Centre's (LPSC) aerospike rocket component, called the module ablation cooling system of the aerospike engine, will be examined. Using off-the-shelf and rolling out of active use steam-propelled ablation mechanisms, the aerospike rocket component deflects significant heat away from the tip of the rocket engine. The engines are valuable as some designs allow the rocket to be more efficiently re-ignited mid-flight, extending mission flexibility and reducing costs. The coolant system is important, as leaks will significantly degrade its overall efficiency. Leakage while in use can result in combustion chamber damage; tubes having holes are considered incomplete, as sleeves are slipped over the intertubes, stitched together, and then attached to rapid-prototyped manifolds. MISUMI-quality pipe fittings are then heat-sinked and brazed onto a monolithic plate using brass-based brazing foil as a filler. These processes are incredibly sensitive to slight modifications in quality. The center operates under NASA Marshall Space Flight; Leaky LFRC intertube leakage results in incomplete modules, but ongoing work is resulting in a clean and balanced final product.

7. Challenges and Limitations of Machine Learning in Aerospace Manufacturing

Challenges & Limitations of ML in Aerospace Manufacturing

Sharp Staff quoted to: Dash et al. (2017) Seven Challenges of Big Data Analytics in Aerospace Manufacturing

Kyle T. Brady, Jeffrey L. Valent, & Qusay H. Mahmoud

To cite this article: Kyle T. Brady, Jeffrey L. Valent, & Qusay H. Mahmoud (2021) Leveraging Machine Learning for Predictive Analytics in U.S. Aerospace Manufacturing: Techniques and Case Studies, Journal of Computing and Information Science in Engineering, 21:3, 031002.

Challenges & Limitations of Machine Learning in Aerospace Manufacturing

Applying machine learning techniques to predict manufacturing outcomes is not without its challenges. Aircraft structure design and certification are deeply rooted and specialized in the aerospace domain. However, the largely publicly available approaches used in this study can be applied to any discrete or batch production process.

Challenges exist accessing real-use case data due to proprietary information. Consequently, relevant data might be limited to testing or non-conforming cases. In addition, relevant data may be of poor quality due to extensive feature engineering. One proposed approach to address relevant feature design might be combining domain knowledge for rule-based models and tree-based feature importance for machine learning models.

Model interpretation approaches have been previously addressed as a limitation in the black-box explanation and human-computer interaction pipelines. Furthermore, the environmental and operational conditions of a large scale and complex system such as an aircraft are difficult to simulate with physical testing alone. Per the economic investment in the aircraft industry, access to the necessary production data for this type of study is highly limited and may be subject to export-controlled information. As a result, many public studies apply similar data-driven-based approaches to the one detailed in this paper using Kolmogorov-Smirnov statistics.

7.1. Data Quality and Availability

Data quality and availability present challenges in applying machine learning in the field of aerospace manufacturing. Access to field-level data and global-scale trends and features were highly dependent on prior working relationships and contracts between researchers and data-providing agencies. Data is expensive to collect, especially high-velocity and high-end data, or suffering from communication latency of more secure, intellectual property data. It was further determined that even though lower-cost sources of data such as process inputs and results could be made available, this data was collected at long time scales, commensurate for human observation and intervention, prohibiting its use in predictive machine learning applications. This section of the effort successfully developed data to support training, validation, and deployment machine learning models.

Utilized data used was collections of raw digital input-to-controller data from aerospace composite-winded area Cincinnati gantry machines. The data was stripped of potential data privacy-related or excessive bandwidth-related information and was collected, cleaned, and stored on the private HPCDC computational infrastructure. This effort used the BReX machine learning testbed as HPCDC for discrete ML tests and the adaptive manufacturing floor for HPCDC-based real-time ML model visualization. Collected data scales varied greatly between datasets. Repeated MacGyver setups of the ML system were used by overcoming computing infrastructure resource challenges of equitable use, independent of ability to pay, and personnel limitations regarding ML software and analytic domain expertise.

7.2. Interpretability of Models

Due to the shortcomings of linear regression in solving business problems in aerospace and high-precision manufacturing, we have gravitated towards automatic learning and more sophisticated modeling techniques. The solution we prefer in general is the random forest regressor, mainly because of its superior predictive performance. However, in using these model-building techniques, one key downside becomes model interpretability. This has consequences for understanding and explaining what predictive models have learned. Once adopted, models are expected to produce accurate predictions on an independent data set. Given the UNdata dataset, which comprises a limited number of observations, the model's accuracy can be measured by cross-validation and a small standard deviation of approximately 10%.

In the context of U.S. aerospace manufacturing, it is necessary to understand how machine learning models work. When predicting, models recognize important feature classifications, which in turn encourages trust among domain experts. Where models are not only used for analytics, but also inform decisions and system controls, model interpretability is key: decision-makers need to know why the model makes predictions, and they might have to make decisions based on them. Consequently, understanding models and embracing their full potential is a function of interpretability. While emphasis is primarily placed on predictive performance, models must also be interpretable enough for validation, verification, and comprehensibility among domain experts. However, in practice, domain and machine learning experts still face the challenge of having to share a common language. Thus, "human"-level learning techniques allow model wrong decisions to be visualized as human auxiliary tools. What is the graphical representation of this printed circuit board? Can multi-class classification be decoded (visualizing this classifier)? What element in a deep neural network should a heat map or function analyzer identify?

8. Future Directions and Emerging Trends

Overall, machine learning is an evolving field and has the potential to create many more business problems in the future. It can positively impact the aerospace and defense industry, particularly for supply chain and production work. Most recent studies have combined advanced machine learning techniques in their approach. Future studies in the aerospace industry are expected to include a better set of methods in combination to improve the future work outcome. For instance, the application of weighted sequential pattern mining in conjunction with Principal Component Analysis and deep learning would be an interesting approach. Integration of spatio-temporal modeling to track ancillary problems in the future looks bright. Also, the most recent trends in the U.S.A involve specific applications of drone production and await future analysis.

There is an extensive market for machine learning applications in the context of 'aerospace manufacturing'. Companies are using developments in predictive modeling to get real-time alerts on potential equipment and component failure. The future trends will be led by research in assistive intelligence, machine learning in drones, the use of GPU in predictive analytics, its application, mechanisms for sequencing network for deep convolutional networks, updates in MSpin algorithm, and more real-world case studies. The future applications for machine learning in aerospace manufacturing will involve the latest computer algorithms. The algorithms for the development of RNN models have to be improved to understand branching and conjugate turning sequences in the assembly of a fighter aircraft. There is a need to understand explainable AI that can provide security in systems using machine learning applications. Machine learning techniques are expected to be used to develop turbulence in drone productions. The use of deep learning techniques to analyze the operational parameters of drones is an emerging domain of research.

8.1. Explainable AI in Aerospace Manufacturing

Explainable AI in aerospace manufacturing. State-of-the-art artificial intelligence techniques, such as deep learning, show a great level of technical performance in various datasets and scenarios. This makes them very attractive in a number of applications, including those in the aerospace manufacturing sector. However, deep learning techniques are associated with several challenges in terms of interpretability. This makes it difficult to utilize the model predictions for decision making in the domain of aerospace manufacturing, where certification authorities demand that machine learning models provide proof of trustworthiness, especially for AI-based active learning. This problem has given rise to a new area called Explainable AI (XAI), which specifically focuses on developing machine learning models for developing explainable predictions.

We believe that the emphasis of our work in explainable AI can be particularly beneficial for machine learning applications in the U.S. aerospace and manufacturing industry. The vision of AI/ML in the aerospace industry is to move from organizational domains where AI models deliver predictions, models supported by experts that deliver decisions, to AI models that deliver decisions that are federated with experts to increase safety and reduce certification barrier to entry for AI applications in aviation. The replacement of domain knowledge in a predictive model supported decision by deep learning has the potential to remove human-based biases and systematic variability from a decision. While this will be useful in certain domains such as automobile sales, it is far from acceptable when you are operating in an environment such as aerospace, which is regulated from sensor to individual components to entire systems up to the global air traffic management network in the sky. This, in turn, creates an innovative use case of AI/ML in aerospace, called Trusted AI, which gives you a range of explainable AI in aerospace, design thinking, and implications.

Reference:

1. Pelluru, Karthik. "Integrate security practices and compliance requirements into DevOps processes." MZ Computing Journal 2.2 (2021): 1-19.
2. Nimmagadda, Venkata Siva Prakash. "AI-Powered Risk Management and Mitigation Strategies in Finance: Advanced Models, Techniques, and Real-World Applications." Journal of Science & Technology 1.1 (2020): 338-383.
3. Machireddy, Jeshwanth Reddy, and Harini Devapatla. "Leveraging Robotic Process Automation (RPA) with AI and Machine Learning for Scalable Data Science Workflows in Cloud-Based Data Warehousing Environments." Australian Journal of Machine Learning Research & Applications 2.2 (2022): 234-261.
4. Potla, Ravi Teja. "Integrating AI and IoT with Salesforce: A Framework for Digital Transformation in the Manufacturing Industry." Journal of Science & Technology 4.1 (2023): 125-135.
5. Singh, Puneet. "Streamlining Telecom Customer Support with AI-Enhanced IVR and Chat." Journal of Artificial Intelligence Research and Applications 3.1 (2023): 443-479.

6. Sreerama, Jeevan, Mahendher Govindasingh Krishnasingh, and Venkatesha Prabhu Rambabu. "Machine Learning for Fraud Detection in Insurance and Retail: Integration Strategies and Implementation." Journal of Artificial Intelligence Research and Applications 2.2 (2022): 205-260.

7. Rambabu, Venkatesha Prabhu, Munivel Devan, and Chandan Jnana Murthy. "Real-Time Data Integration in Retail: Improving Supply Chain and Customer Experience." Journal of Computational Intelligence and Robotics 3.1 (2023): 85-122.

8. Althati, Chandrashekar, Venkatesha Prabhu Rambabu, and Munivel Devan. "Big Data Integration in the Insurance Industry: Enhancing Underwriting and Fraud Detection." Journal of Computational Intelligence and Robotics 3.1 (2023): 123-162.

9. Krothapalli, Bhavani, Lavanya Shanmugam, and Jim Todd Sunder Singh. "Streamlining Operations: A Comparative Analysis of Enterprise Integration Strategies in the Insurance and Retail Industries." Journal of Science & Technology 2.3 (2021): 93-144.

10. Amsa Selvaraj, Priya Ranjan Parida, and Chandan Jnana Murthy, “AI/ML-Based Entity Recognition from Images for Parsing Information from US Driver’s Licenses and Paychecks”, Journal of AI-Assisted Scientific Discovery, vol. 3, no. 1, pp. 475–515, May 2023

11. Deepak Venkatachalam, Pradeep Manivannan, and Jim Todd Sunder Singh, “Enhancing Retail Customer Experience through MarTech Solutions: A Case Study of Nordstrom”, J. Sci. Tech., vol. 3, no. 5, pp. 12–47, Sep. 2022

12. Pradeep Manivannan, Deepak Venkatachalam, and Priya Ranjan Parida, “Building and Maintaining Robust Data Architectures for Effective Data-Driven Marketing Campaigns and Personalization”, Australian Journal of Machine Learning Research & Applications, vol. 1, no. 2, pp. 168–208, Dec. 2021

13. Praveen Sivathapandi, Priya Ranjan Parida, and Chandan Jnana Murthy. “Transforming Automotive Telematics With AI/ML: Data Analysis, Predictive Maintenance, and Enhanced Vehicle Performance”. Journal of Science & Technology, vol. 4, no. 4, Aug. 2023, pp. 85-127

14. Priya Ranjan Parida, Jim Todd Sunder Singh, and Amsa Selvaraj, "Real-Time Automated Anomaly Detection in Microservices Using Advanced AI/ML Techniques", J. of Artificial Int. Research and App., vol. 3, no. 1, pp. 514–545, Apr. 2023

15. Sharmila Ramasundaram Sudharsanam, Pradeep Manivannan, and Deepak Venkatachalam. "Strategic Analysis of High Conversion Ratios from Marketing Qualified Leads to Sales Qualified Leads in B2B Campaigns: A Case Study on High MQL-to-SQL Ratios". Journal of Science & Technology, vol. 2, no. 2, Apr. 2021, pp. 231-269

16. Jasrotia, Manojdeep Singh. "Unlocking Efficiency: A Comprehensive Approach to Lean In-Plant Logistics." *International Journal of Science and Research (IJSR)* 13.3 (2024): 1579-1587.

17. Gayam, Swaroop Reddy. "AI-Driven Customer Support in E-Commerce: Advanced Techniques for Chatbots, Virtual Assistants, and Sentiment Analysis." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 92-123.

18. Nimmagadda, Venkata Siva Prakash. "AI-Powered Predictive Analytics for Retail Supply Chain Risk Management: Advanced Techniques, Applications, and Real-World Case Studies." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 152-194.

19. Putha, Sudharshan. "AI-Driven Energy Management in Manufacturing: Optimizing Energy Consumption and Reducing Operational Costs." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 313-353.

20. Sahu, Mohit Kumar. "Machine Learning for Anti-Money Laundering (AML) in Banking: Advanced Techniques, Models, and Real-World Case Studies." Journal of Science & Technology 1.1 (2020): 384-424.

21. Kasaraneni, Bhavani Prasad. "Advanced Artificial Intelligence Techniques for Predictive Analytics in Life Insurance: Enhancing Risk Assessment and Pricing Accuracy." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 547-588.

22. Kondapaka, Krishna Kanth. "Advanced AI Techniques for Optimizing Claims Management in Insurance: Models, Applications, and Real-World Case Studies." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 637-668.

23. Kasaraneni, Ramana Kumar. "AI-Enhanced Cybersecurity in Smart Manufacturing: Protecting Industrial Control Systems from Cyber Threats." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 747-784.

24. Pattyam, Sandeep Pushyamitra. "AI in Data Science for Healthcare: Advanced Techniques for Disease Prediction, Treatment Optimization, and Patient Management." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 417-455.

25. Kuna, Siva Sarana. "AI-Powered Solutions for Automated Customer Support in Life Insurance: Techniques, Tools, and Real-World Applications." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 529-560.

26. Sengottaiyan, Krishnamoorthy, and Manojdeep Singh Jasrotia. "SLP (Systematic Layout Planning) for Enhanced Plant Layout Efficiency." International Journal of Science and Research (IJSR) 13.6 (2024): 820-827.

27. Gayam, Swaroop Reddy. "AI-Driven Fraud Detection in E-Commerce: Advanced Techniques for Anomaly Detection, Transaction Monitoring, and Risk Mitigation." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 124-151.

28. Nimmagadda, Venkata Siva Prakash. "AI-Powered Risk Assessment Models in Property and Casualty Insurance: Techniques, Applications, and Real-World Case Studies." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 194-226.

29. Putha, Sudharshan. "AI-Driven Metabolomics: Uncovering Metabolic Pathways and Biomarkers for Disease Diagnosis and Treatment." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 354-391.

30. Sahu, Mohit Kumar. "AI-Based Supply Chain Optimization in Manufacturing: Enhancing Demand Forecasting and Inventory Management." Journal of Science & Technology 1.1 (2020): 424-464.

31. Kasaraneni, Bhavani Prasad. "Advanced Machine Learning Algorithms for Loss Prediction in Property Insurance: Techniques and Real-World Applications." Journal of Science & Technology 1.1 (2020): 553-597.

32. Kondapaka, Krishna Kanth. "Advanced AI Techniques for Retail Supply Chain Sustainability: Models, Applications, and Real-World Case Studies." Journal of Science & Technology 1.1 (2020): 636-669.

33. Kasaraneni, Ramana Kumar. "AI-Enhanced Energy Management Systems for Electric Vehicles: Optimizing Battery Performance and Longevity." Journal of Science & Technology 1.1 (2020): 670-708.

34. Pattyam, Sandeep Pushyamitra. "AI in Data Science for Predictive Analytics: Techniques for Model Development, Validation, and Deployment." Journal of Science & Technology 1.1 (2020): 511-552.

35. Kuna, Siva Sarana. "AI-Powered Solutions for Automated Underwriting in Auto Insurance: Techniques, Tools, and Best Practices." Journal of Science & Technology 1.1 (2020): 597-636.

36. Selvaraj, Akila, Deepak Venkatachalam, and Jim Todd Sunder Singh. "Advanced Telematics and Real-Time Data Analytics in the Automotive Industry: Leveraging Edge Computing for Predictive Vehicle Maintenance and Performance Optimization." Journal of Artificial Intelligence Research and Applications 3.1 (2023): 581-622.
37. Selvaraj, Amsa, Debasish Paul, and Rajalakshmi Soundarapandiyan. "Synthetic Data for Customer Behavior Analysis in Financial Services: Leveraging AI/ML to Model and Predict Consumer Financial Actions." Journal of Artificial Intelligence Research 2.2 (2022): 218-258.
38. Paul, Debasish, Rajalakshmi Soundarapandiyan, and Gowrisankar Krishnamoorthy. "Security-First Approaches to CI/CD in Cloud-Computing Platforms: Enhancing DevSecOps Practices." Australian Journal of Machine Learning Research & Applications 1.1 (2021): 184-225.
39. Venkatachalam, Deepak, Jeevan Sreeram, and Rajalakshmi Soundarapandiyan. "Large Language Models in Retail: Best Practices for Training, Personalization, and Real-Time

Customer Interaction in E-Commerce Platforms." Journal of Artificial Intelligence Research and Applications 4.1 (2024): 539-592.

40. Namperumal, Gunaseelan, Rajalakshmi Soundarapandiyan, and Priya Ranjan Parida. "Cloud-Driven Human Capital Management Solutions: A Comprehensive Analysis of Scalability, Security, and Compliance in Global Enterprises." Australian Journal of Machine Learning Research & Applications 2.2 (2022): 501-549.
41. Kurkute, Mahadu Vinayak, Gunaseelan Namperumal, and Akila Selvaraj. "Scalable Development and Deployment of LLMs in Manufacturing: Leveraging AI to Enhance Predictive Maintenance, Quality Control, and Process Automation." Australian Journal of Machine Learning Research & Applications 3.2 (2023): 381-430.
42. Soundarapandiyan, Rajalakshmi, Deepak Venkatachalam, and Akila Selvaraj. "Real-Time Data Analytics in Connected Vehicles: Enhancing Telematics Systems for Autonomous Driving and Intelligent Transportation Systems." Australian Journal of Machine Learning Research & Applications 3.1 (2023): 420-461.
43. Sivathapandi, Praveen, Venkatesha Prabhu Rambabu, and Yeswanth Surampudi. "Advanced CI/CD Pipelines in Multi-Tenant Cloud Platforms: Strategies for Secure and Efficient Deployment." Journal of Science & Technology 2.4 (2021): 212-252.
44. Sudharsanam, Sharmila Ramasundaram, Gunaseelan Namperumal, and Akila Selvaraj. "Integrating AI/ML Workloads with Serverless Cloud Computing: Optimizing Cost and Performance for Dynamic, Event-Driven Applications." Journal of Science & Technology 3.3 (2022): 286-325.