**Process Optimization in Biomanufacturing**

**Alexandr M. Sokolov and Karnik Aswani**

Arkansas State University

State University, AR 72467, USA

[asokolov@astate.edu](mailto:asokolov@astate.edu), karnik.aswani@smail.astate.edu

**Abstract**

**Purpose –** Biomanufacturing is a distinctive method utilized in the manufacturing industry to produce biological products including microorganisms. There are potential errors and issues in optimizing the biological models and processes to create these specialized products. Biological models range from mathematical optimization models to model-based dynamical systems. This study will analyze and provide potential strategies to increase the efficiency and accuracy of these biological models.

**Design/methods approach –** Using machine learning models and experimentations, several biological parameters were improved to increase the accuracy, efficiency, and reduce the costs of these biological models. New sensing technologies combined with software models show promising results such as outcome improvement, prediction, smart process control, and fingerprinting.

**Findings –** The proposed machine learning models and experiments were executed and analyzed. Optimizing process variables would improve the processes of creating these biological products and the outcome predictions. Nutrient utilization and system productivity would improve the overall performance of the processes conducted. According to the findings, product quality control had to be enhanced to reduce the number of errors and issues in the experimental processes.

**Originality/value –** Experimental processes were conducted in research and development centers. WuXi Biologics Ultra-high productivity platform technology is used for continuous biomanufacturing processes to deliver the processes.

**Keywords**

Biomanufacturing; optimization processes; product quality; sensing; advanced control

**1. Introduction**

Biomanufacturing is an extraordinary method to create biological products using analytical techniques and extracting biomolecules including properties and functions from organisms. The environmental conditions where these biological processes will be executed require a high efficiency rate and low manufacturing costs. Efficiency and costs can be analyzed using analytical techniques such as artificial intelligence and machine learning techniques to implement strategies to improve the quality of the processes. In this case, process optimization would be the most effective and feasible method for consideration.

**2. Literature Review**

2.1 Applications of Biomanufacturing

Biomanufacturing organisms are used in manufacturing to create biological products and processes, however, the production and cost need to be low. Process optimization is a highly valuable method in microbial manufacturing to boost microbial manufacturing performance and economic benefit. Manufacturing of biological products requires a biomanufacturing system that consists of unit operations including upstreaming and down streaming for production and purification (Zheng, 2022). It would require a trial-and-error method to analyze the microbial metabolism and fermentation performance from biological parts such as molecules, cells, and the microenvironment (Wan, 2023). Viruses such as the recombinant adeno-associated virus (rAAV) have been used in upstream processing for medicinal therapies, however, the costs and production were high. Researchers resolved this by improving the critical process parameters using soft-sensing technologies combined with predictive modeling (Iglesias,2023). Another microorganism named microalgae has been used to operate biological processes more efficiently using on-line sensing for monitoring biological parameters (Havlik,2022). The on-line sensing technologies are software sensors, mathematical process models, fuzzy logic, and artificial neural networks. A similar case study was conducted related to microalgae using genome-scale models and controllers in a controlled environment where microalgae would perform photosynthesis (Li, 2023). However, this study was mainly to maximize productivity while minimizing nutrient costs.

2.2 Growth of Biological Products

The place where these biological products can be created are biofoundries which are highly automated facilities that create and test the cycle of the manufacturing process of these microorganisms (Tellechea-Luzardo, 2022). However, at the same, it is very time-consuming and expensive. Researchers were able to resolve this issue through process planning, simulation, and optimization of the processes. Additionally, a technology called WuXi Biologics’ Ultra-High Productivity platform (WuXiUP) is used in the biomanufacturing processes to improve its efficiency using several process parameters (Zhou, 2021).

2.3 Bacteria in Biomanufacturing

A bacteria-inhibitory substance was produced and studied to test if it counters bacteria such as *Listeria monoctyogenes*. The device called the Brain Heart Infusion was utilized to grow and expand the number of bacteria; however, the composition of the substance was needed to improve the production of the bacteria-inhibitory substance (Jawan, 2021). The best solution was modeling the optimization process to identify improvements in the composition of the substance. A similar study was conducted using monoclonal antibodies produced by GS-NS0 cells. In this study, a mathematical model was developed to predict the dynamic of the nutrient uptake and intracellular ATP content to quantify and characterize the energy metabolism in cells (Quiroga-Campano, 2018).

2.4 Alternative Methods to Process Parameters

Process parameters are not the only methods that can improve biological processes in terms of efficiency and cost. Dynamical systems are another method that can describe the system of biological processes using physical and mathematical properties (Xu, 2023). In addition, sensors consist of numerous key variables that were used in process characterization and outcome predictions to optimize process performance. This method would be highly applicable to potential applications such as fingerprinting or smart process control in the quality improvement of systems and processes (Reyes,2022). A linear programming model was utilized in the manufacturing industries to optimize multiple biopharmaceutical products to meet consumer demands. The result was a greater optimality with lower time and manufacturing costs (Siganporia, 2013). Control strategies were developed to minimize the amount of energy and material consumption in producing the biological products. Two of the control strategies are a semicontinuous purification process and a cell culture system in a bioreactor (Papathanasiou, 2017).

2.5 Challenges in Biomanufacturing

Despite the benefits of biomanufacturing, biomanufacturing does face significant challenges that can affect biological processes (Xie, 2021). Significant challenges would include complexity, high variability, long lead time, and very limited process data. To tackle these challenges for any microorganism in any setting would require using statistical methods such as probabilistic knowledge graphs and game theory to facilitate the production processes in terms of quality-by-design and stability control. Another method that would apply to tackle these challenges is the Monte Carlo method which can determine the process uncertainty among the process parameters by incorporating the process inputs into a deterministic model to simulate more possible scenarios for the outputs (Furcht,2022). Additionally, the work that the researchers have done was on adjusting the sizing units for the operations to improve the utilization rate and reduce manufacturing costs. From this study, researchers were able to provide more insights into the outcomes without the probability of any occurrences. The probability of occurrences is not the only issue in these experiments, contaminants can also affect the manufacturing processes. To solve this problem, researchers suggested the installation of process analytical technology within the production processes to measure the critical quality parameters including screening for potential biological contaminations (Janghorban,2023).

**3. Methodology/Framework of Study**

* 1. Problem

Optimizing the performance of penicillin and drug production raises the question of how the production time of the process can be reduced while improving resource efficiency. Determining the factors that affect the controlled environment where penicillin and drug grows is crucial. This can include temperature, pH, nutrient utilization, pressure, light exposure, and other necessary factors.

* 1. Exploratory Data Analysis

Gathering the basic statistics of the Penicillin and drug production datasets is essential as it provides a deeper understanding of the data and how it can contribute to potential machine-learning models. The most effective techniques that would be used are correlation tests and creating scatter plots to determine the relationship between two variables. From these correlation tests and creating the scatter plots, it would tell which variables are highly correlated.

* 1. Feature Selection

Picking the most applicable feature to create the machine learning models. There is a possibility that new features will be created that can be implemented into the machine learning models to get better results. The variables that will be utilized in the Penicillin dataset would be the amount of penicillin produced at the beginning and end of the batch process, and the total yield of the overall batch process. The variables that will be utilized in the drug production dataset would be biomass, protein production, impurity, and purity.

* 1. Model Selection

Determining the appropriate machine learning models is crucial in terms of evaluating the accuracy of these models. The most appropriate machine learning models would be linear regression, logistic regression, neural networks, density-based clustering, lasso regression, and ridge regression. Linear regression will be used to calculate the correlation among the variables including visualizing the relationships. In addition to the linear regression, a neural network will be conducted along with the linear regression model to determine if this model is efficient in predicting the amount of protein being produced. Logistic regression will be used for binary classification of the fault in the Penicillin dataset. A lasso regression will be used to determine if multicollinearity is presented in the dataset or not. Ridge regression is like the lasso regression as well, however, both models will be conducted to double-check if multicollinearity exists or not. Lastly, a density-based clustering model will be conducted to determine which batches of the Penicillin and drug are closer together.

* 1. Model Training

Training the machine learning models would involve splitting the dataset including the feature variables and a target variable. This method is applied to logistic regression, lasso regression, ridge regression, and neural networks. The density-based clustering will be trained by inputting the number of clusters with sample data to mix the actual data with the sample data. In this way, the accuracy score will be determined by training how well it classifies the actual data from the sample data. The linear regression will not be trained since its main purpose is to determine the strength of the relationship between two variables.

* 1. Model Optimization

Optimizing the machine learning models would include changing the parameters to increase the accuracy score among the models. This can include changing the number of clusters, the threshold to make the model work, inputting the number of sample data, and other potential changes. The neural networks would be the most potential challenge in optimizing the model since it involves many layers in the network to be adjusted just to achieve the highest accuracy score possible.

* 1. Interpretation and Documentation

Interpreting and documenting the model results can be used for future research involving similar cases. The most appropriate way to document the results would be to create a formal scientific report where scientists and engineers can read and analyze where improvements and adjustments need to be made in the biomanufacturing processes and models. In this paper, no scientific report will be created, however, the interpretation and findings of the results will be analyzed and concluded to inform potential scientists and engineers of potential recommendations that can utilized in the facilities where these processes take place.

**4. Case Study**

4.1 Benefits of Biomanufacturing

Biomanufacturing in microorganisms is making an incredible impact on single-use technologies in the future. The economic benefits of single-use technologies would be lower capital investments by % 40 and increased profits from biological products. In addition, production costs will be lower if industries utilize single-use technologies for biomanufacturing. The benefit of lower production costs would be a higher savings rate to either implement new processes or improve the current processes. Also, the quality of the products would increase.

4.2 Single-Use Technologies

In addition, the five most important reasons for utilizing single-use technologies are the elimination of cleaning requirements, reduced risk of cross-contamination, faster turnaround between campaigns, and increased convenience and flexibility of disposable technologies. On a related note, the single-use technology can couple with the process analytical technology to improve the monitoring of the manufacturing processes, facilitate the controls, and make any necessary corrections during the campaign of the turnover. Another area that can be expanded with single-use technology would be microbial fermentation which can provide enough mass and heat for the fermentation process to work.

**5. Analysis**

5.1 Correlation Test

Three correlation tests were conducted in the Penicillin datasets, and five correlation tests were conducted in the drug production dataset. Three variables for the Penicilin dataset were utilized in the correlation tests which are Penicllin\_yield\_total..kg., Penicllin\_harvested\_during\_batch.kg., and Penicllin\_harvested\_end\_of\_batch..Kg. Four variables for the drug production were utilized in the correlation tests which are Initial Biomass X\_0, Protein X\_F, Impurity I\_F, and Purity.

For the Penicillin dataset, the first correlation was between the Penicillin harvested during the process and the Penicillin harvested at the end of the process showed a numeric value of 0.4977216. The second correlation test between the total yield of Penicillin and Penicillin harvested during the process showed a numeric value of 0.7408194. The third correlation test between the total yield of Penicillin and the Penicillin harvested end of the process showed a numeric value of 0.9513157.

For the drug production dataset, the first correlation was between the initial biomass of the drug and the protein production showed a numeric value of 0.8566366. The second correlation test between the initial biomass of the drug and the drug impurity showed a numeric value of 0.8427418. The third correlation test between the initial biomass and the drug purity showed a numeric value of -0.3611378. The fourth correlation test between protein production and drug purity showed a numeric value of -0.2997708. The fifth correlation test between protein production and drug impurity showed a numeric value of 0.9880507.

Table 1. Correlation Tests for Penicillin Production

|  |  |
| --- | --- |
| Penicillin harvested during the process vs Penicillin harvested at the end of the process | 0.4977216 |
| Total yield of Penicillin vs Penicillin harvested during the process | 0.7408194 |
| Total yield of Penicillin vs Penicillin harvested at the end of the process | 0.9513157 |

Table 2. Correlation Tests for Drug Production

|  |  |
| --- | --- |
| Initial biomass vs protein production | 0.8566366 |
| Initial biomass vs drug impurity | 0.8427418 |
| Initial Biomass vs drug purity | -0.3611378 |
| Protein production vs drug purity | -0.2997708 |
| Protein production vs drug impurity | 0.9880507 |

5.2 Machine Learning Models

The R squares and the accuracy score are some of the most important statistical factors to compare plus it provides whether the model is a good fit to use or not for future processes. The higher the accuracy score, the better the model does a great job of either classifying the data or making predictions. The higher the r square, the higher the variance of the data. Other statistical factors are mentioned in this analysis such as the distance between clusters and coefficients for each variable which are crucial factors for determining the fitness of the model.

Table 3. R Square for Machine Learning Models

|  |  |
| --- | --- |
| Logistic Regression | 0.1152281 |
| Lasso Regression for Penicillin Production | 0.9991480 |
| Lasso Regression for Drug Production | 0.9938379 |
| Ridge Regression for Penicillin Production | 0.9954719 |
| Ridge Regression for Drug Production | 0.9806922 |

Table 4. Accuracy Score for Machine Learning Models

|  |  |
| --- | --- |
| Logistic Regression | 0.8888888 |
| Neural Networks | 0.8990650 |

5.3 Linear Regression

Linear regression was conducted in both Penicillin and drug production datasets. Three linear regression models were conducted in the Penicillin dataset while five linear regression models were conducted in the drug production dataset.

A graph of a graph with blue dots

Description automatically generated with medium confidence

Figure 1. Linear Regression for Harvested During Batch vs Harvested End of Batch. It shows there is an upward trend and most of the data points are close to the curve.

The first linear regression model in the Pencillin dataset is harvested during batch vs harvested end of batch. It shows a moderate positive relationship in an upward curve until it starts to decrease midway. This implies it starts to increase gradually at an optimal temperature meaning the highest temperature where the Penicillin starts to grow.

A graph with blue dots

Description automatically generated

Figure 2. Linear Regression for Total Yield vs Harvested End of Batch. It shows a nearly perfect straight line with most of the data points either on or near the line.

The second linear regression model is the total yield versus the harvest at the end of the batch. It shows a perfect linear regression line indicating a very strong positive relationship. This implies that as one sample of the batch is being harvested at the end of the batch, the total yield increases gradually over time until it reaches its maximum amount.

A graph with blue dots

Description automatically generated

Figure 3. Linear Regression for Total Yield vs Harvested During the Batch. It shows a nearly perfect straight line with most of the data points either on or near the line.

The third linear regression model is the total yield versus the harvest during the batch. It shows a nearly perfect linear regression line indicating a moderate positive relationship. This implies that as one sample of the batch is being harvested at the beginning, the total yield increases over time until it reaches its maximum amount.

A graph with blue dots

Description automatically generated

Figure 4. Linear Regression for Biomass vs Protein Production. It shows most of the data points either on or close to the upward slope.

The fourth linear regression model in the drug production dataset is Biomass vs Protein Production. It shows a linear regression line indicating a positive relationship. This implies that as one sample of the protein is produced, the biomass increases over time.

A graph of a line graph

Description automatically generated with medium confidence

Figure 5. Linear Regression for Biomass vs Impurity. It shows most of the data points either on or close to the upward slope.

The fifth linear regression model in the drug production dataset is Biomass vs Impurity. It shows a linear regression line indicating a positive relationship. This implies that as the impurity of the protein is mixed with the protein component, the biomass increases over time.

A graph of a mass plot

Description automatically generated with medium confidence

Figure 6. Linear Regression for Biomass vs Purity. It shows some of the data points either on or close to the downward slope.

The sixth linear regression model in the drug production dataset is Biomass vs Purity. It shows a downward trend at midway then it starts to increase afterward. This implies that as the purity of the protein is mixed with the protein component, the biomass decreases over time.

A graph of a protein production

Description automatically generated

Figure 7. Linear Regression for Protein Production vs Purity. It shows a few of the data points either on or close to the downward slope.

The seventh linear regression model in the drug production dataset is Protein Production vs Purity. It shows a linear regression in the downward direction indicating a negative relationship. This implies that as one sample of the protein is produced, the purity decreases over time.

A graph of a protein production

Description automatically generated

Figure 8. Linear Regression for Protein Production vs Impurity. It shows most of the data points either on or close to the upward slope.

The eighth linear regression model in the drug production dataset is Protein Production vs Impurity. It shows a linear regression line indicating a positive relationship. This implies that as one sample of the protein is produced, the impurity decreases over time.

5.4 Logistic Regression

A logistic regression was conducted in this analysis for mainly the Fault.ref.0.NoFault.1.Fault. variable in the Pencilin dataset. The analysis was conducted using only 1s and 0s. The variable importance, value of each predictor variable, for the logistic regression was 1.871951 for Penicllin\_harvested\_during\_batch.kg., 1.872092 for Penicllin\_harvested\_end\_of\_batch..kg., and 1.871995 for Penicllin\_yield\_total..kg. The variance inflation factor, value to determine the multicollinearity of the model, is 49030230 for Penicllin\_harvested\_during\_batch.kg., 200876286 for Penicllin\_harvested\_end\_of\_batch..kg., and 358833636 for Penicllin\_yield\_total..kg. In logistic regression, there is no such as an actual R2 instead there is McFadden’s R2 which is a predictive technique for calculating the R2. The McFadden’s R2 for the logistic regression is 0.1152281. The accuracy score for this regression was 0.888888888888889.

5.5 Density-Based Clustering

A density-based clustering method was conducted in this analysis to determine the similarity among potential different clusters within the drug production dataset. The sum of squared distance of the samples within each cluster is 2462.401 while the average similarity of samples within clusters is 0.5080485.

A graph of a graph

Description automatically generated with medium confidence

Figure 9. Density-Based Clustering for Drug Production. Four clusters ranging from tiny to large in this diagram.

5.6 Lasso Regression

A lasso regression model was conducted both in the drug and Penicillin datasets. The variables utilized in the Penicillin dataset are Penicllin\_yield\_total..kg, Penicllin\_harvested\_during\_batch.kg.,andPenicllin\_harvested\_end\_of\_batch..kg. The R-squared of this model was 0.999148 meaning % 99.91. For the drug production dataset, the variables are Initial Biomass X\_0, Protein X\_F, Impurity I\_F, and Purity. The R-squared of this model was 0.9938379 meaning % 99.38. Additionally, the coefficients are 0.6928379 for Initial Biomass X\_0, 0.3013727 for Impurity I\_F, and 92.6573065 for Purity.

A graph of a function

Description automatically generated with medium confidence

Figure 10. Lasso Regression for Drug Production. It shows a gradual increase representing an exponential function meaning fewer predictors are approaching zero coefficients.

A graph of a function

Description automatically generated

Figure 11. Lasso Regression for Penicillin Production. It shows a gradual increase representing an exponential function meaning fewer predictors are approaching zero coefficients.

5.7 Ridge Regression

A ridge regression model was conducted both in the drug and Penicillin datasets. The variables utilized in the Penicillin dataset are Penicllin\_yield\_total..kg, Penicllin\_harvested\_during\_batch.kg.,andPenicllin\_harvested\_end\_of\_batch..kg. The R-squared of this model was 0.9954719 meaning % 99.54. For the drug production dataset, the variables are Initial Biomass X\_0, Protein X\_F, Impurity I\_F, and Purity. The R-squared of this model was 0.9806922 meaning % 98.07. Additionally, the coefficients are 0.8268772 for Initial Biomass X\_0, 0.2979440 for Impurity I\_F, and 96.9434743 for Purity.

A graph of a line

Description automatically generated

Figure 12. Ridge Regression for Drug Production. It shows a gradual increase representing an exponential function meaning fewer predictors are approaching zero coefficients.

A graph of a function

Description automatically generated

Figure 13. Ridge Regression for Penicillin Production. It shows a gradual increase representing an exponential function meaning fewer predictors are approaching zero coefficients.

5.8 Neural Networks

A neural network model was conducted in the drug dataset. The variables utilized in the drug dataset are Initial Biomass X\_0, Protein X\_F, Impurity I\_F, and Purity. The accuracy score of this model was 0.8990651.

A diagram of a network

Description automatically generated

Figure 14. Neural Networks for Drug Production. It shows different sets of layers for classifying the protein as its outcome.

**6. Results Interpretation**

6.1 Correlation Tests

For the Penicillin dataset, the total yield of Penicillin and the Penicillin harvested end of the process correlation test shows a strong positive relationship indicating the two variables are directly proportional to each other. This means as the Penicillin harvested at the end of the process increases, so does the total yield of Penicillin. The Penicillin harvested during the process and the end process correlation test has a moderate positive relationship indicating the two variables are directly proportional to each other with external factors affecting the relationship. The total yield of the total yield of Penicillin and the Penicillin harvested during the process correlation test shows a slightly strong positive relationship indicating the two variables are directly proportional to each other with potential external factors affecting the relationship.

6.2 Logistic Regression for Penicillin Production

Logistic regression is not a great model for determining if there is a contamination issue in the Penicillin production process including the predictive power. McFadden’s R2 indicates a low predictive power, and the model does not fit very well with the data. On the other hand, the accuracy score is very high which can predict the contamination issue accurately. Overall, this model can potentially predict the contamination issue, however, there are potential issues that need to be addressed such as adding or changing the parameters including rewriting the whole code to make this model reliable.

6.2 Ridge Regression for Penicillin Production

The ridge regression is a great model for predicting the total yield production of Penicillin with a high R2 numeric value. It represents a significant variance in the Penicillin being harvested using several of the predictor variables conducted in the analysis. Additionally, it provides insights into the relationships between the total yield production and process rate. Additionally, it can predict the Penicillin yield, quality, and other outputs as well, however in this case, it would predict the yield being harvested after the end of the batch in an upward slope. Also, it can show the relationships among the variables to determine the quality of the antibiotics being produced. This proves that the ridge regression model for Penicillin production is accurate, however, it can be improved by adjusting some of the parameters such as the amount of sample.

6.3 Ridge Regression for Drug Production

The ridge regression is a great model for predicting drug production with a high R2 numeric value. It represents a significant variance in the drug being produced using several of the predictor variables conducted in the analysis. Additionally, it can predict the drug yield, quality, and other outputs as well, however in this case, it would predict the yield being produced after the end of the process in an upward slope. Also, it can show the relationships among the variables to determine the quality of the drugs being produced. This proves that the ridge regression model for drug production is accurate, however, it can be improved by adjusting some of the parameters such as the amount of sample.

6.4 Lasso Regression for Penicillin Production

The lasso regression is a great model for predicting the total yield production of Penicillin with a high R2 numeric value. Similarly, to the ridge regression, it does a great job of representing the variance and providing insights into the relationships of the Penicillin being harvested. Additionally, it can predict the Penicillin yield, quality, and other outputs as well. Similarly, it can determine the quality of the antibiotics being produced. It proves that the lasso regression model for antibiotic production is accurate, however, it can be improved by adjusting some of the parameters such as the amount of sample.

6.5 Lasso Regression for Drug Production

The lasso regression is a great model for predicting drug production with a high R2 numeric value. Similarly, to the ridge regression, it does a great job of representing the variance and providing insights into the relationships of the drugs being produced. Additionally, it can predict the drug yield, quality, and other outputs as well. Similarly, it can determine the quality of the drug being produced. It proves that the lasso regression model for drug production is accurate, however, it can be improved by adjusting some of the parameters such as the amount of sample.

6.6 Neural Networks for Drug Production

The third model great for predicting drug production with a high R2 numeric value is neural networks. The accuracy rate is high like the other models, and it does a great job of predicting drug production by classifying different properties of the drugs with the process variables provided.

6.7 Density-Based Clustering for Drug Production

The density-based clustering is a great model for determining the distance between the clusters and the number of clusters specifically for drug production. The numeric values suggest a moderate clustering meaning the clusters are related but not significantly related. The different sizes of the clusters suggest some clusters are very closely related and some are not based on the properties of the drugs being produced.

**7. Validation**

7.1 Potential Biases

Despite the high R2 and accuracy score from all the machine learning models conducted in this analysis, there are potential biases that affect the models. One possible bias is not accurately measuring the amount of Penicillin and dose of the drug in the harvesting process including the start and end. Another bias would be the temperature where chemical reactions can affect the production rate in terms of product quality, yield, and process efficiency. The third bias is pressure which can change the product properties leading to imbalance or incomplete biological reactions. The fourth bias is acidic levels typically using the pH scale which involves changing the acidic levels to alter the product of the production process. The fifth bias would be nutrient utilization changing the level of nutrients in the Penicillin batch during and before the harvesting process. The sixth bias would be not using the correct amount of solution concentrate in the process which can affect the productivity rate (Coffman, 2021).

7.2 Potential Solutions to Counter Biases

All the biases stated can potentially increase the accuracy score of each model such as by adding more hyperparameters, adjusting current hyperparameters, or rewriting the entire code with a few extra features. The better method is to choose other potential machine learning models such as k-nearest neighbors, decision trees, discriminant analysis, etc. Another method is complex; however, it can work, it is cleaning the data further such as changing the data types or scaling the data further. Changing the data types can be used for classification instead of regression models for better classifying the different biological properties and a potentially higher accuracy score associated with drug production. Scaling the data further for both datasets can train and test the actual data better with a shorter response time creating a higher accuracy score.

**8. Discussion from Managerial Perspective, Industry Perspective, and/or Academic Perspective**

8.1 Handling of Optimization Processes

Projects involving optimization processes require easy computerization such as machine learning models to predict the amount of Penicillin and drugs being produced. The cost needs to be reduced specifically operating costs for these optimization processes to move forward. The capability of handling these optimization processes is crucial as well since it requires the right resources, the right equipment, and the right people. The ease of use and flexibility are potential challenges since creating biological products takes a huge amount of time and effort. Also, investment in these optimization processes is crucial as well cause these optimization processes again take a lot of time and effort which requires a high cost. If the return on investment is lower than expected, then these biopharmaceutical facilities will start to reduce costs in other areas of these facilities to make up for the losses (Cotts, 2014).

8.2 Issues in Optimization Processes

All these optimization processes are in large biopharmaceutical facilities which require high costs, risks, and time trade-offs if one option in the plan of the optimization process fails. Several problems in optimization processes can include incorrect capacity planning, under-capacity, and high-profile company acquisitions. Additionally, some biopharmaceutical companies are considering the risks involved in these optimization processes such as failed batches, natural disasters, incorrect market demand forecasts, or clinical trial failure (Siganporia,2013).

8.3 Maintenance of Biofoundries

In a general sense, facilities require people including employees who can speed up the optimization processes if these employees are equipped with the right tools, right skills, and amount of experience (Cotts, 2014). By hiring the right people, these optimizations can produce more biological products to gain more revenue in the future. Also, on a related note, maintenance is crucial in operating these optimization processes which would require people as well. This can include occasional inspection of the facility where the optimization processes are being conducted. The main reason is to prevent any potential failures such as machinery, malfunctioning equipment, and systematic errors which can affect the production of biological products in terms of components and functions. Cleaning the rooms is crucial because potential hazards such as spills, wastes, and other harmful contaminants can affect the productivity rate, including failed experiments (Cotts, 2014).

8.4 Efficiency Improvement

Efficiency has been the most important metric in the analysis and what strategies could be implemented to improve the efficiency? Three strategies can be implemented to improve the efficiency of the optimization processes. One strategy would be focusing on planning proactive work which is to plan what tasks and equipment are needed and completed at first before executing the other minor tasks. The second strategy would be to provide the proper tools that are needed for performing these optimization processes. The main reason would be the wrong equipment even if it is related to the actual equipment can potentially cause failures and errors. The third strategy would be setting up clear expectations with creative and logical thinking among team members including supervision roles as well (Cotts, 2014).

8.5 Other Potential Risks

Despite its benefits, there are potential risks involved in the life cycle of creating these biological products even with nearly perfect adjustments to the technology and processes. Risks would include failure to maintain the components and processes and not gathering enough accurate data, (Meredith, 2022).

**9. Recommendations**

9.1 Techniques

Several techniques are advised to improve the efficiency in the optimization process of converting the biological components into usable biological products. One technique would be using connectors in a bioprocess system that is well validated, strong, easy to use, and highly reliable (Mirasol, 2017). This technique would be most applicable to protein production as it involves protein components in producing the drugs.

9.2 Upgrading Systems in Biofoundries

There is a possibility that some of the systems in biofoundries such as the HVAC can affect the pressure in the optimization process, so the best solution is to update the HVAC to its fullest capacity when the optimization processes are occurring specifically at the time of the day (Mirasol, 2017). Another potential system that can improve the efficiency of the optimization processes is the microprocessor lighting control system which can provide an optimal light level to prevent these biological components from being ruptured or destroyed (Cotts, 2014).

9.3 Automation in Biofoundries

In biofoundries, people are responsible for these optimization processes, however, many people are not required to execute these optimization processes. One potential technique is to use automation that can handle more of the complex optimization processes that produce the most demanded products. This can include vaccines, cell therapies, and more biological products. The three biggest benefits of automation would be increasing the productivity rate, decreasing operating costs, and increasing accountability. The productivity rate will increase with fewer people, however, as mentioned earlier, it would require the right people with the right skills. A decrease in operating costs would increase the accuracy of the number of biological products being produced. Accountability needs to make sure the data being collected and imported is accurate so that processes are not flawed and potentially increase the costs of the optimization processes (Cotts, 2014).

9.4 Preventive Maintenance

Preventive maintenance is the first recommendation because scheduling these optimization processes is necessary since any process, equipment, or machine needs to be repaired or replaced to prevent hazards or disasters in the facilities. A biopharmaceutical facility would be replacing worn-out equipment that can potentially harm the samples being used to create these biological products. Computation resources need to be updated regularly to prevent a potential crash in the computing system that can lose all progress of the processes being executed. Since process optimization requires gathering data, it is essential to regularly update the data management software which could include PhpMyAdmin, Oracle, and other data management tools (Tellechea-Luzardo, 2022). In addition, it can lower costs and improve the efficiency of the optimization processes. This can include adjusting the temperature to meet its optimal temperature, balancing the chemical reactions, and changing the level of nutrient levels (Cotts,2014).

9.5 Corrective Maintenance

Corrective maintenance is another recommendation for the maintenance of the optimization processes. The main reason is that there is a potential possibility that the number of ingredients in making the drugs, or the antibiotics will either be low or high depending on how many samples were utilized and the system's failure to quantify the amount. If the batches are low, then some of the batches fail which requires an immediate removal of these batches, if possible, to prevent the failed batches from mixing in with grown batches. If the batches are high, then the production process is increasing the supply which can potentially exceed the demand. The excess demand would cause a loss in profits, so the best way is to carefully measure and quantify the samples close to the exact amount for the production. This would include using machine learning models, resetting equipment, and restoring machines to an optimal level of performance (Cotts,2014).

9.6 Predictive Maintenance

Predictive maintenance is the third recommendation for the monitoring of the optimization processes. The main reason is that there is a possibility that potential failure will occur from a variety of reasons such as inaccurate data, malfunction of equipment or machines, and other potential failures. One way is to catch the problems earlier before the equipment or machines fail such as the bioreactors, stirred tanks, etc. This approach can be difficult since these processes are complex. Another way is to repair equipment and machines that have not been utilized too much even if the time for these processes is delayed, so in that way, more products will be produced in the long run (Cotts,2014).

9.7 Using Sensors to Improve Results

Another great recommendation is to use appropriate sensors to improve the efficiency and accuracy of the processes occurring in the facilities. All the sensors recommended should have the following characteristics which are selectivity, reproducibility, accuracy, stability, sensitivity, resolution, linearity, response time, and robustness. The selectivity of biological components is crucial which in this case are the samples of the Penicillin and proteins being produced. Reproducibility can reproduce the correct number of samples in the previous experiments to generate identical outputs. The accuracy was conducted earlier in the analysis; however, these sensors can improve the accuracy further. The stability of these sensors can make sure the exact outputs are being produced from the previous experiments which is like reproducibility, however, stability is just maintaining the processes throughout the whole time. Resolution can create small changes to the process variables in both optimization processes to improve efficiency. Linearity is used to measure the outputs in terms of its expected measurement from the actual measurement. The response time plays a significant role as well since production needs to be faster with lower time which is the main purpose of optimization of these processes. This would require a change in the input variables to speed up the experiments to achieve the most accurate results. Robustness plays a significant role as well cause environmental conditions such as a dirty room, wastes, contaminants, or unused substances from previous experiments can affect the usability of these sensors (Reyes, 2022).

9.8 Optical Sensors Through the Use of Electricity

Optical sensors are sensors that can monitor the oxygen, pH of the substances, and dissolved carbon dioxide with the use of electricity through the lens. The biggest advantage of this sensor would be to detect parameter shifts such as if the variables start to decrease or increase that can affect the quality of the products being produced. The equipment that would be most affected by these sensors is the bioreactors where oxygen and carbon dioxide are dissolved along with the pH of organic compounds (Reyes, 2022).

9.9 Mass Spectrometry to Analyze Data

Mass spectrometry is a powerful analytical tool that can monitor carbon dioxide, oxygen, volatile organic compounds, and other chemical substances. Since it is an analytical tool, gathering data is necessary to perform quality control tests to find correlations among the variables which is like the variables analyzed earlier, however, an addition to the analysis is feedback control to ensure the variables apply to the processes. If it is not applicable, rearrangement of the mathematical models to conduct the processes is necessary and more concentration will be utilized on the variables that affect the processes the most (Reyes, 2022).

9.9 Electrochemical Sensors to Monitor Chemical Reactions

Electrochemical sensors are sensors that can monitor the chemical reactions from electrical properties and charged substances occurring in the optimization processes. The two most important parameters to monitor are the pH and dissolved oxygen. These parameters are important to measure to make the products being produced in a stable environment, however, the drawback is that the dissolved oxygen will last for a temporary period. It is highly advisable to use these sensors if the products are being produced too low or the accuracy score from the machine learning models is too low (Reyes, 2022).

9.10 Potential Causes from Human Errors

Human errors account for some of the potential failures in these optimization processes, so the best solution is to train the employees to reduce the errors, however, it would depend on the budget of the facility. If there is a strict budget, then there needs to be some adjustments in reducing the cost of other parts in the facilities and fulfilling in other areas where the optimization processes occur. The other costs in the optimization processes would be raw materials, labor, quality control, and purification (Zheng, 2022).

9.11 Biopharma 4.0 for Production Optimization

Another way to reduce human error is using Biopharma 4.0 which are digital technology including artificial intelligence and machine learning to optimize biopharmaceutical production efficiency by minimizing human intervention. Biopharma 4.0 consists of soft sensors that can reduce time and production costs, generate data, improve the control of bioprocesses, increase product quality, and reduce raw material waste without the use of many humans. However, a few humans would be needed to maintain and repair the soft sensors in case of any potential failures (Iglesias, 2023).

9.12 Potential Solutions to Lower Risk

Risks are involved in these optimization processes including a strict budget. To avoid a strict budget or potentially reduce it is to measure and quantify how much costs are being utilized in the machine, equipment, processes, workers, and other factors as well. Some of the most important inputs to consider when assessing these risks are the labor and material cause those two resources can use up most of the spending. In complex optimization processes, reducing the number of people and keeping the most qualified people can reduce labor costs significantly. The best way to reduce the costs of materials is to pick out the equipment and machines that are going to be utilized the most and create fewer errors while the processes are being conducted. In addition, figuring out the best way to reduce costs for optimizing processes can reduce labor and material costs (Meredith, 2022).

9.13 Best Recommendations

Overall, all these recommendations would reduce time and increase the efficiency of production while reducing the number of errors in processing these optimizations. Other strategies and recommendations can improve the efficiency and reduce the time of these optimization processes; however, these recommendations can provide the best accurate results in the end.

**10. Conclusion**

Biomanufacturing is a unique biological technique that extracts biological components to create biological products. One important factor in making these biological products is optimizing processes that require machines, equipment, qualified people, and process strategies. Efficiency and time are some of the most important factors that affect the optimization processes. In terms of process strategies, artificial intelligence, and machine learning have become popular data analysis tools for investigating the potential causes of producing fewer biological products. Some of the machine learning techniques utilized in this analysis were linear regression, logistic regression, neural networks, lasso regression, ridge regression, density-based clustering, and correlation tests.

Two datasets were utilized which are the Penicillin and drug production to analyze the trend of the production process and predict the next cycle of production. The first step was the correlation tests to determine the relationships among several variables. Three positive relationships were determined in both datasets while two negative relationships were determined in the drug production dataset only. Linear regression was conducted for both datasets to demonstrate the relationships plus determine the strength of these relationships by closely observing how well the data points are on or near the regression line. For the Penicillin dataset, logistic regression was conducted to determine the number of faults in the harvesting process of Penicillin. Neural networks were conducted in the drug dataset along with linear regression to optimize the process control for parameter adjustments. Lasso and ridge regression were conducted in both datasets to predict the outcomes of the products being produced. Density-based clustering was conducted in the drug production dataset to categorize the different groups of drugs with similar properties and monitor the process performance over time.

After all the machine learning models were created, the results were documented to compare all the accuracy scores to determine whether all the machine learning models could predict and classify Penicillin and drug production. Once the results were documented, potential biases were discussed as to how the bias affected the process performance of the Penicillin and the drug being produced. Afterwards, there was a discussion of how all the machine-learning models can be used in industries, and what project management strategies are used in process optimization in biomanufacturing. Finally, recommendations were discussed on how to improve efficiency and reduce the time in process optimization further down the line as manufacturing biological products is becoming more important in the next decade or so.

**11. Relevance to Engineering Management**

11.1 Importance of Process Optimization

Process optimization in biomanufacturing is relevant to engineering management because it can aid engineers and scientists in making the byproducts of biological products in a more efficient and faster way. Biologists and biological engineers create biological products including antibodies, cell therapies, etc. Additionally, optimizing processes would require production planning to determine the size of the facility in terms of cost, risk, and time trade-offs (Siganporia, 2013).

11.2 Production Planning

Production planning is widely concerned with mathematical models including linear programming; however, machine learning models can still play a significant role as well. The construction of a biomanufacturing facility would require lead time; in other words, how long will it take for the facility to build? Commission would play a role as well in terms of a contract where the project manager and contractor will want a particular process to be done for the public. Validation is crucial as well since the projects that are being implemented in these facilities are complex and need to be validated before allowing the facilities to distribute the products to the public. Another concern is that production planning is focused on batch rather than continuous processes. Two main reasons for this concern are time and cost. The third reason is to predict the capacity of the biological products being produced which can help the business decision of determining whether outsourcing production is appropriate or not depending on the market (Siganporia, 2013).

11.3 Cell Therapies

Cell therapies are important for the medical field as they can treat various diseases including cancer, neurodegenerative disorders, and autoimmune disorders. Manufacturing cell therapies would require the use of stirred tank reactors, liquid chromatography systems, and cross-filtration technologies. All equipment and systems require maintenance of the technologies and systems including material and cost and reducing the amount of time due to high demand from the public. Additionally, another concern about producing cell therapies is the facility that needs to be manufactured within a specific period which can raise costs, so maintenance and optimizing processes can reallocate the costs for a process to other costs that need to be covered in the facility. This can be achieved by the strategies that were mentioned earlier for the analysis of Penicillin and drug production (Aijaz, 2018).

11.4 Antibodies

Antibodies are another important biological product in the medical field which involves project management expertise such as product quality, and reduction in manual handling to increase safety and efficiency of manufacturing processes. The quality of antibodies is very crucial as it can increase the productivity rate and volume as it is being administrated. Additionally, this can involve different states of the antibody being produced plus the charge isoform, aggregates, and the different organic compounds linked to the chemical reactions (Karst,2016).

11.5 Skills Required for Optimization Processes

Another reason why optimizing processes in the biomanufacturing industry is relevant to engineering management because the main reason is that taking care of optimizing processes requires project management, engineering, technical, and soft skills. Process optimization in the biomanufacturing industry is considered a complex project as it involves qualified people to operate these processes, many components including machinery and non-machinery, tons of cost, and lots of time.

**References**

Aijaz, A., Li, M., Smith, D., Khong, D., LeBlon, C., Fenton, O. S., Olabisi, R. M., Libutti, S., Tischfield, J., Maus,

M. V., Deans, R., Barcia, R. N., Anderson, D. G., Ritz, J., Preti, R., & Parekkadan, B. (2018).

Biomanufacturing for Clinically Advanced Cell Therapies. *Nature Biomedical Engineering*, *2*(6), 362–376.

<https://doi.org/10.1038/s41551-018-0246-6>

Coffman, J., Bibbo, K., Brower, M., Forbes, R., Guros, N., Horowski, B., Lu, R., Mahajan, R., Patil, U., Rose, S., &

Shultz, J. (2021). The design basis for the integrated and continuous biomanufacturing framework. *Biotechnology and Bioengineering*, *118*(9), 3323–3333. https://doi.org/10.1002/bit.27697

Cotts, David G., et al. *The Facility Management Handbook*. American Management Association, 2014.

Furcht, Christopher, et al. “Use of Monte Carlo Simulations for Improved Facility Fit Planning in Downstream Biomanufacturing and Technology Transfer.” *Biotechnology Progress*, vol. 39, no. 1, Jan. 2023, pp. 1–10. *EBSCOhost*, https://doi-org.ezproxy.library.astate.edu/10.1002/btpr.3306.

Grogan, Michael. “Neuralnet: Train and Test Neural Networks Using r: R-Bloggers.” *R*, 10 Oct. 2018, www.r-bloggers.com/2018/10/neuralnet-train-and-test-neural-networks-using-r/.

Havlik, Ivo, et al. “On-line monitoring of biological parameters in microalgal bioprocesses using optical methods.” *Energies*, vol. 15, no. 3, 25 Jan. 2022, p. 875, https://doi.org/10.3390/en15030875.

Iglesias, Cristovão Freitas, et al. “Raav Manufacturing: The challenges of soft sensing during upstream processing.” *Bioengineering*, vol. 10, no. 2, 8 Feb. 2023, p. 229, https://doi.org/10.3390/bioengineering10020229.

Janghorban, Mohammad, et al. “Methods and analysis of biological contaminants in the biomanufacturing industry.” *Chemosensors*, vol. 11, no. 5, 18 May 2023, p. 298, https://doi.org/10.3390/chemosensors11050298.

Jawan, Roslina, et al. “Evaluation of the estimation capability of response surface methodology and artificial neural network for the optimization of bacteriocin-like inhibitory substances production by Lactococcus Lactis GH1.” *Microorganisms*, vol. 9, no. 3, 12 Mar. 2021, p. 579, https://doi.org/10.3390/microorganisms9030579.

Karst, Daniel J., et al. “Process performance and product quality in an integrated continuous antibody production process.” *Biotechnology and Bioengineering*, vol. 114, no. 2, 17 Aug. 2016, pp. 298–307, https://doi.org/10.1002/bit.26069.

Li, Chien-Ting, et al. “Optimization of nutrient utilization efficiency and productivity for algal cultures under light and dark cycles using genome-scale model process control.” *Npj Systems Biology and Applications*, vol. 9, no. 1, 15 Mar. 2023, https://doi.org/10.1038/s41540-022-00260-7.

Meredith, Jack R., and Scott M. Shafer. *Project Management a Strategic Managerial Approach*. Wiley, 2022.

Mirasol, Feliza; “Connector Integrity in Single-Use Biomanufacturing Systems.” *BioPharm International*, MJH Life

Sciences, Nov. 2017, [www.biopharminternational.com/view/connector-integrity-single-use-biomanufacturing-systems](http://www.biopharminternational.com/view/connector-integrity-single-use-biomanufacturing-systems).

Papathanasiou, Maria M., et al. “Advanced model‐based control strategies for the intensification of upstream and downstream processing in Mab Production.” *Biotechnology Progress*, vol. 33, no. 4, 26 Apr. 2017, pp. 966–988, https://doi.org/10.1002/btpr.2483.

Quiroga-Campano, A. L., Panoskaltsis, N., & Mantalaris, A. (2018). Energy-based culture medium design for

biomanufacturing optimization: A case study in monoclonal antibody production by GS-NS0 cells. *Metabolic Engineering*, *47*, 21–30. <https://doi.org/10.1016/j.ymben.2018.02.013>

Reyes, Sebastian Juan, et al. “Modern sensor tools and techniques for monitoring, controlling, and improving cell culture processes.” *Processes*, vol. 10, no. 2, 18 Jan. 2022, p. 189, <https://doi.org/10.3390/pr10020189>.

Siganporia, Cyrus C., et al. “Capacity Planning for batch and perfusion bioprocesses across multiple biopharmaceutical facilities.” *Biotechnology Progress*, vol. 30, no. 3, 24 Jan. 2014, pp. 594–606, https://doi.org/10.1002/btpr.1860.

Tellechea-Luzardo, Jonathan, et al. “Fast biofoundries: Coping with the challenges of biomanufacturing.” *Trends in Biotechnology*, vol. 40, no. 7, July 2022, pp. 831–842, https://doi.org/10.1016/j.tibtech.2021.12.006.

Wan, Shengtong, et al. “Current Advances for OMICS-Guided Process Optimization of Microbial Manufacturing - Bioresources and Bioprocessing.” *SpringerOpen*, Springer NatureSingapore, 30 Apr. 2023, bioresourcesbioprocessing.springeropen.com/articles/10.1186/s40643023-00647-2.

Xie, W., Wang, B., Li, C., Xie, D., & Auclair, J. (2021). Interpretable biomanufacturing process risk and sensitivity analyses for quality‐by‐design and Stability Control. *Naval Research Logistics (NRL)*, *69*(3), 461–483. <https://doi.org/10.1002/nav.22019>

Xu, Gongxian, and Zijia Liu. “Sequential Geometric Programming Method for Parameter Estimation of a Nonlinear System in Microbial Continuous Fermentation.” *International Journal of Chemical Engineering*, Hindawi, 10 Oct. 2023, www.hindawi.com/journals/ijce/2023/8072920/.

Zach, Bobbitt. “Ridge Regression in R (Step-by-Step).” *Statology*, 13 Nov. 2020, www.statology.org/ridge-regression-in-r/.

Zach, Bobbitt. “Lasso Regression in R (Step-by-Step).” *Statology*, 13 Nov. 2020, www.statology.org/lasso-regression-in-r/.

Zach, Bobbitt. “How to Perform Simple Linear Regression in R (Step-by-Step).” *Statology*, 26 Oct. 2020,

[www.statology.org/simple-linear-regression-in-r/](http://www.statology.org/simple-linear-regression-in-r/).

Zach, Bobbitt. “How to Perform Logistic Regression in R (Step-by-Step).” *Statology*, 29 Sept. 2021,

[www.statology.org/logistic-regression-in-r/](http://www.statology.org/logistic-regression-in-r/).

Zach, Bobbitt. “How to Perform a Correlation Test in R (with Examples).” *Statology*, 22 Nov. 2021,

[www.statology.org/correlation-test-in-r/](http://www.statology.org/correlation-test-in-r/).

Zheng, Hua, et al. “Policy optimization in Dynamic Bayesian network hybrid models of biomanufacturing processes.” *INFORMS Journal on Computing*, vol. 35, no. 1, Jan. 2023, pp. 66–82, <https://doi.or> g/10.1287/ijoc.2022.1232.

Zhou, Hang, et al. “Improving an intensified and integrated continuous bioprocess platform for biologics manufacturing.” *Biotechnology and Bioengineering*, vol. 118, no. 9, 14 Apr. 2021, pp. 3618–3623, https://doi.org/10.1002/bit.27768.

**About the authors**

**Alexandr M. Sokolov, Ph.D.,** is a faculty in the Engineering Management and Construction Management Department of the College of Engineering and Computer Science at Arkansas State University. He holds a B.S., where he focused on Bioinformatics from the University of Tennessee Knoxville, an M.B.A., in Finance from Lincoln Memorial University, and a Ph.D., in Industrial Systems Engineering, Engineering Management from the University of Tennessee Space Institute. His teaching experience includes multiple institutions dealing with Engineering, Construction, Management, and Technology disciplines. He is focusing on research dealing with Engineering Management, Performance Management, and Interdisciplinary Studies.

**Karnik Aswani, MSDS**, is a graduate student pursuing an MS in Engineering Management at the Department of the College of Engineering and Computer Science at Arkansas State University. He holds a BA in Economics from the University of Nevada, Las Vegas, and an MS in Data Science from Eastern University where he applied his economic and data science tools to real-life scenarios. Currently, he is pursuing his third MS in Applied Economics to advance his skills in economics. Additionally, he has done tons of internships and volunteer work in business, project management, economics, and data science.