BMT22
4th International Conference
Business Meets Technology
Ansbach, 7th – 9th July 2022

PATTERN RECOGNITION PROGRAMMING TO PREDICT PRODUCTIVITY OF YARROWIA LIPOLYTICA DSM 3286 FOR CITRIC ACID PRODUCTION.

Hain, Christopher[a] and Gaisser, Sibylle[a]

[a]*Ansbach University of Applied Sciences, Germany*

ABSTRACT: The non-conventional yeast Yarrowia lipolytica is attracting increasing attention due to its potential to produce large amounts of organic acids from hydrophobic substrates. Due to the steadily increasing demand for citric acid in the industrial sector, the aim of this scientific work was to develop a predictive model of the citric acid productivity of the strain Yarrowia lipolytica DSM3286. As a basis for this, the optical density, pH, cell number and citric acid were determined in 18 identical mixtures.

The citric acid concentration (mean values of the measured concentration over time) follows a linear increase. Based on this, the mathematical calculation operation of linear regression was selected for modeling the prediction model in Python. The following coefficients were determined for the variables used in the learning algorithm:

- time: $6{,}104 * 10^{-4}$
- OD: $-1{,}224 * 10^{-1}$
- pH value: $-4{,}043 * 10^{-1}$
- Cell count: $1{,}749 * 10^{-8}$

In final validation of the program, a result accuracy of 86.5% was obtained. The result obtained in the present scientific work shows that by means of simple linear regression, over a cultivation period of 13 days, a prediction of the citric acid productivity of strain Yarrowia lipolytica DSM3286 is possible.

Keywords: *Yarrowia lipolytica; Citric acid; Predictive model; Python*

How to cite: Hain, Christopher and Gaisser, Sibylle. 2022. Pattern recognition programming to predict productivity of Yarrowia lipolytica DSM 3286 for citric acid production. In Proc.: 4th International Conference Business Meets Technology. Ansbach, 7th – 9th July 2022.

1. INTRODUCTION

Yarrowia lipolytica belongs taxonomically to the ascomycetes and is an obligate aerobic, apatho-genic, heterothallic and oleaginous yeast (Fickers et al., 2002; Barth et al, 2003). Fat- and protein-rich substrates represent the natural habitat (Barth et al. 2003), but the obligate aerobic yeast *Y. lipolytica* can also grow on substrates such as glucose, fructose, ethanol, or acetate.

Y. lipolytica is of biotechnological interest mainly because under certain culture conditions the apathogenic strain can secrete various organic acids, such as citric acid or isocitric acid, into the medium in high concentrations of up to 200 g* L^{-1}. Due to these properties and the simple procedural cultivation of the strain, it represents an ideal microorganism for the biotechnological production of citric acid (Gonçalves et al. 2014). Citric acid is the most consumed organic acid in the world and is mainly used as an acidifier, preservative, and antioxidant in various industries (Morgunov et al., 2018).

More and more scientists are also turning to self-learning algorithms and artificial intelligence (AI) in the natural sciences to do their work more efficiently and effectively. The use of self-learning algorithms and artificial intelligence to optimize processes is therefore becoming increasingly important in the natural sciences. Python is one of the most popular languages for scientific programming. Python is an object-oriented, open-source, universally interpretable programming language that is well suited for standard programming tasks. (Millman und Aivazis 2011; Dubois 2007)

Based on the previous studies on citric acid production in *Y. lipolytica* DSM 3286, the present scientific work aimed to create a Python programming to predict productivity from a total of 18 identical cultivation approaches. For this purpose, the data of optical density, pH, cell number and measured citric acid concentration were first documented and evaluated at regular intervals in each batch. This was followed by the programming of the learning algorithm in Python, the preparation of the training data sets and the evaluation of the program.

2. METHOD

The test execution as well as the programming was done as shown within the following flowcharts.

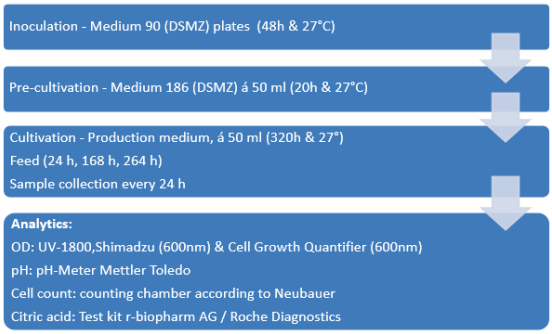


Figure 1: Execution of the experiment shown as a flowchart

Multiple linear regression was used to program the citric acid prediction in Python and implemented as follows.

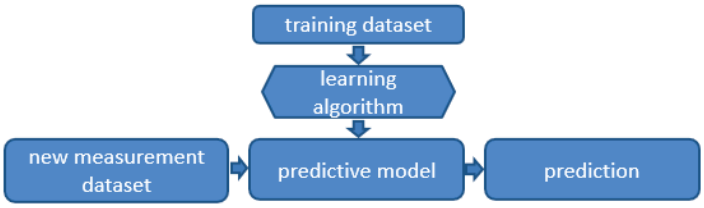
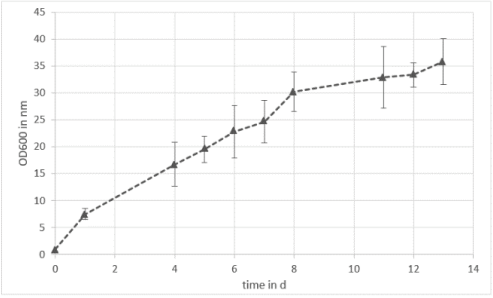


Figure 2: The learning algorithm is programmed from the training data. Via the input of new measurement data into the prediction model, the prediction of the unknown value takes place. (Raschka 2017)

3. RESULTS

For comparability, the mean values from the respective samples with the same sampling times were calculated and illustrated in the following graphs.



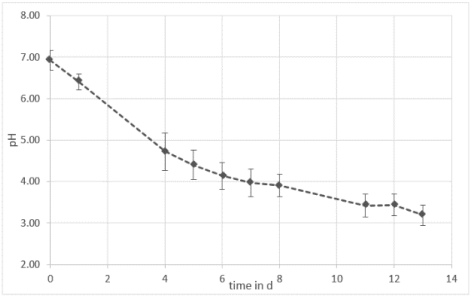
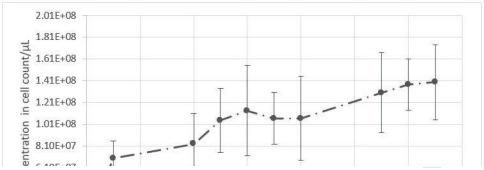


Figure 3: The average optical density and its standard deviation as well as the average pH value and its standard deviation over the cultivation period of 13 days; N=18.

Figure 4: The average cell number/µL with standard deviation as well as the average citric acid yield with standard deviation over the cultivation period of 13 days; N=18.

The following figures show the programming results.



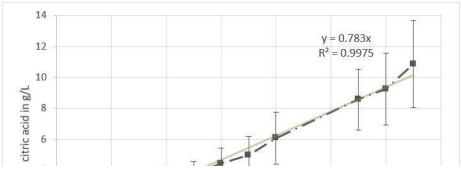
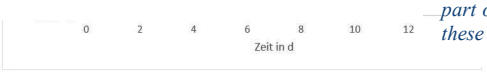
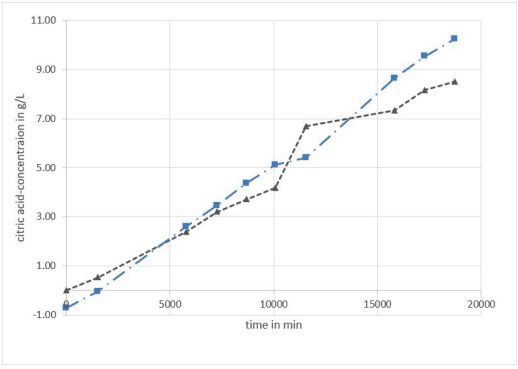


Figure 6: Multiple Linear Regression source code used, which was developed in Python.

*Figure 5: Course of measured (triangle/ grey) and predicted (quadrilateral/ blue) citric acid concentration of batch 7.1 (no part of the training data sets). The average result accuracy of these two runs is **92%**.*




```python
#Multiple Regression

import pandas as pd

from sklearn import linear_model

data = pd.read_csv("./excelfiles/Datensatz1-10.csv")
x = data[['mins', 'OD', 'pH', 'cells']]
y = data['cs']

regr = linear_model.LinearRegression()
regr.fit(x, y)

#Probeneingabe:
#predictedCS = regr.predict([[Zeit in min,OD,pH-Wert,Zellzahl]])
predictedCS = regr.predict([[1440,8.18,6.39,76000000]])

print("Citronensäure-Konzentration: ")
print(predictedCS)
print("----------------")
print("Koeffizienten: ")
print(regr.coef_)
print("----------------")
print("Ergebnisgenauigkeit:")
print(regr.score(x,y))
print("----------------")
```



4. CONCLUSION

Considering the result accuracy of the programmed model of 86.5 %, this can be rated as good compared to other biological models, which show a result accuracy of close to 90 %.(An, 2019). To increase the accuracy, more test data sets should be incorporated into the model. Considering the exclusion of individual variables and the resulting accuracy of results, it is recommended to measure all four variables in future approaches and to include them in the program. In conclusion, the model for predicting citric acid production in strain Y. lipolytica works within the cultivation period of 13 days.

REFERENCES

An, J.; You Z.; (2019). Sequence-based Prediction of Protein-Protein Interactions Using Gray Wolf Opti-mizer–Based Relevance Vector Machine.

Barth, G.; Beckerich, J. M.; Dominguez, A.; Kerscher, S.; Ogrydziak, D.; Titorenko, V.; Gaillardin, C., (2003). Functional genetics of Yarrowia lipolytica. In: de Winde, J. H. und S. Hohmann, Functional Gene-tics of Industrial Yeasts.

Dubois, P. F. (2007): Python: Batteries Included,. In: Computing in Science & Engineering (9).

Fickers, P.; Le Dall, M. T.; Gaillardin, C.; Thonart, P.; Nicaud, J. M., (2003). New disruption cassettes for rapid gene disruption and marker rescue in the yeast Yarrowia lipolytica.

Gonçalves, F. A. G.; Colen, G.; Takahashi, J. A. (2014): Yarrowia lipolytica and its multiple applications in the biotechnological industry. In: TheScientificWorldJournal 2014, S. 476207.

Millman, K. J.; Aivazis, M., (2011). Python for Scientists and Engineers. In: Comput. Sci. Eng. 13 (2), S. 9–12. DOI: 10.1109/MCSE.2011.36.

Morgunov, I. G.; Kamzolova, S; Lunina, J. N.; (2018). Citric Acid Production by Yarrowia lipolytica Yeast on Different Renewable Raw Materials.

Raschka, S.: Machine Learning mit Python. Das Praxis-Handbuch für Data Science, Predictive Analytics und Deep Learning. 1. Auflage. Frechen: MITP (mitp Professional). Available online at https://www.ciando.com/img/books/extract/3958454240_lp.pdf, 2017, last tested on 01.08.2021.