Jurnal Mantik
Volume 4, Number 1, May 2020, pp. 936-943 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks

Willi Akbar Satria[1], Iskandar Fitri[2], Sari Ningsih[3]

Universitas Nasional, Jalan Sawo Manila, Pasar Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520

Email: willysatria208@gmail.com, tektel2001@yahoo.com, sariningsih.lectures@yahoo.com

ARTICLEINFO

Article history:
Received: 04/04/2020
Revised: 20/4/2020
Accepted: 30/5/2020

Keywords:
Artificial Neural Network, Churn Prediction, Deep Learning, Machine Learning, Hard, TensorFlow.

ABSTRACT

The Artificial Neural Network (ANN) is an architect in deep learning inspired by how the human brain works. As the name suggests, this technique works like a biological neural network possessed by a living creature, by accepting input data, processing it in a node/neuron, and displaying it in the output. ANN works by utilizing the large number of layers and the number of existing nodes/neurons to perform tasks, such as performing feature extraction, pattern recognition, regression, and classification. In this journal ANN's architect was used in the design of deep learning model, to solve classification problem, in case of banking churn prediction. The number of layers (layers) used by 4 layers, namely input layer, 2 hidden layer, and output layer with the number of nodes/neurons in sequence as much as 10.6, 6, 1. The final result is a model that is already (ready) deployed into a web-based application that can predict the churn (banking) with an accuracy of 84% on the customer's input data

1. Introduction

Churn rate is the percentage of customers of a service that does not continue to subscribe to a company or stop using products / services offered within a certain period and can affect the profits that can be achieved by the company, no exception for companies engaged in finance, such as the banking industry [1].

Various types of Machine Learning have been developed to predict customer churn rates, and in general, these models have proven their effectiveness in processing and making predictions based on a given dataset [2].

This journal explains the application of Deep Learning to predict churn rates in the banking industry, whether customers will stop or continue to use the services that banks offer. The model is designed using Artificial Neural Network (ANN) [3], which is used to study and process bank customer data and can then be applied to make new predictions that can be used by companies to determine the extent of churn that may occur in the future, so that various losses can be anticipated by making appropriate decisions based on the results of the predicted models [4].

2. Base Theory

A. Machine Learning

Machine Learning (ML) or in Indonesian is called machine learning is a branch of artificial intelligence that learns how computers learn from data [5]. The term machine learning itself is quite confusing or misleading because it has almost no connection with any machine (unless implemented on a machine or robot). The machine here refers to an algorithm or program running on a computer [6]. Unlike conventional algorithms, ML algorithms study correlations between data to obtain

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

information without relying on specific instructions [7]. Other terms that are commonly used are Data Mining, Pattern Recognition, or Knowledge Discovery [8].

Machine Learning can do that by using tools to study data and solve problems based on existing data. ML learning methods are divided into 2, namely:

a) Supervised Learning, is a ML learning technique, by training the model by providing expected data input and output. In this way ML will learn how to work with data to obtain the information needed [9].
b) Unsupervised Learning, in contrast to supervised learning, this technique works by finding hidden patterns or internal structures in data, which will then be used to draw conclusions, based on these data [10].

B. Deep Learning

Deep Learning is one area of machine learning that isinspired by how the human brain works. The human brain consists of billions of neurons that are interconnected to identify patterns and classify different types of information, as well as the way the DL algorithm works [11]. Withutilize the large number of nodes / neurons and layers (layers) in neural networks (or artificial neural networks) that are used for information processing, such as feature extraction, pattern recognition, regression, and classification [12].

C. Neural Network

As the name suggests this technique works like the biological neural network that living things have, by receiving data input, processing it in neurons, and displaying it at the output [13].

Neural Network consists of several layers (layers), as illustrated in Figure 1. Below are the layers:

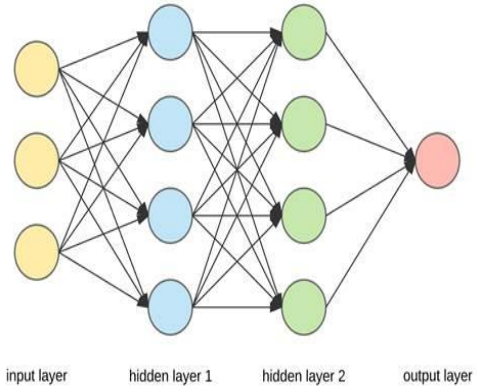

Fig 1. Neural network

a) Input Layer:
The first layer where input data (input) is received, and processed to the next layer.
b) Hidden Layer:
The layer where the data has been received is processed, by performing mathematical calculations. The number of hidden layers is one factor that determines how accurate the output is. Neural networks with several hidden layers in them are usually called deep neural networks.
c) Output Layer:
The last layer that produces output and where the results of data processing are stored.

Neural networks work with a technique called forward propagation and backpropagation, here are the details:

a) Forward Propagation :

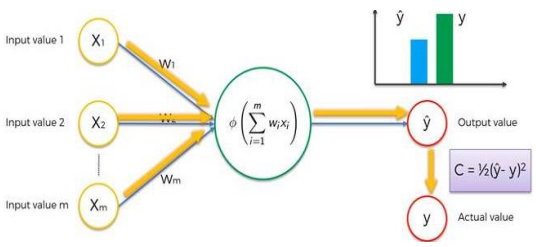

Fig 2. Forward propagation

Figure 2: Forward propagation is done when the model receives input (x) from the input layer in the form of data, which is then operated with a variable weight (w) contained in each connection between nodes /

Accredited "Rank 4" (Sprott 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

neurons (synapses) and then operated with an activation function to generate the predicted output (), which is then compared to the actual output (y) using the loss function / cost function to produce loss. $\hat{y}$

a) Backpropagation:

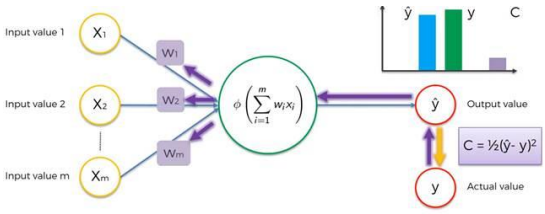

Fig 3. Backpropagation

Figure 3: Backpropagation is done after a loss is detected which is then used by the model to improve predicted output () by adjusting variable weight (w). Both of these processes are commonly called epochs, and are repeated until the model gets optimal output results.$\hat{y}$

D. Tools used

There are several types of tools used to build deep learning model designs, such as, frameworks (backend and frontend), IDE (Integrated Development Environment), and some library software from programming languages (Python). Some of these tools include:

a) Hard

Hard is a software framework (frontend) that functions to design deep learning architectures such as neural networks with the Python programming language. Hard was developed to enable the design of deep learning models to be easier, faster, and without involving a lot of code, making it possible to be used as a tool supporting experiments and prototyping [14].

Hard was developed as part of the ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System) research project and was released in 2018 as an open source under the MIT license.

Keras functions as a frontend framework, which means that it is only a tool for designing DL models, Keras requires frameworks such as Tensorlow or Theano to compile the model.

b) Tensorlow

Tensorflow is a software framework (backend) that has functions to build and deploy machine learning or deep learning models, in contrast to ordinary program models, which require programmers to write program instructions explicitly, with machine learning models, there is no need for instructions on specific programs but by providing training data so that the ML algorithm can study data without having to be explicitly programmed which will later be used to solve the given problem [15].

Tensorflow was developed by Google, with the aim of becoming a common platform in building machine learning. First released in 2015 to the public open source under the Apache Open Source Software License.

Tensorflow has the flexibility so that it can be used to build various types of ML models, but tensorflow is more often used to build Deep Learning with a neural network architecture. Neural networks in tensorflow are used in various fields such as image recognition, speech recognition, image style transfer, or language translation.

3. Design

The Deep Learning model built for this case was built using Keras as a frontend framework and TensorFlow as a backend. The machine learning method used in this model is the supervised learning model, which is one of the categories of machine learning that is designed by training the model by providing training data that has been equipped with labels / attributes that indicate the expected classification or output of the data. The IDE used to build the model is Spyder 3, as well as using the Python 3 programming language.

The design of Artificial Neural Network (ANN) in the deep learning model has several stages, including preparing data, designing models, training models, testing models, implementing models into applications. Following are the details of these stages.

Accredited "Rank 4" (Sprott 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

A. Prepare Data (Data Preprocessing)

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProdu	HasCrCard	IsActiveMemb	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.57	1

Fig 4. Dataset

Figure 4 The data to be used is data in the form of a dataset (with the extension .csv), with a total of 10,000 records and 12 attributes (variables), each representing data from each customer (individual). These attributes consist of, customer id, surename, credit score, geography (categorical), gender (categorical), age, tenure, balance, number of products, has credit cards (categorical), is active member (categorical), estimated salary, and exited which functions as the label of the dataset (0 means false, and 1 means true).

a) Specifies the Attribute to be used as a Feature

From 12 existing attributes, selection is needed to determine what attributes are considered influential in determining the label (prediction results). In this case, 10 attributes will be used as features, namely credit score, geography, gender (categorical), age, tenure, balance, number of products, has credit cards (categorical), is active member (categorical), and estimated salary.

b) Import Data and Preprocessing Data

Name	Type	Size	Value
X	object	(10000, 10)	ndarray object of numpy module
dataset	DataFrame	(10000, 14)	Column names: RowNumber, CustomerId, Surname, CreditScore, Geography, ...
y	int64	(10000,)	[1 0 1 ... 1 1 0]

Fig 5. Results of dataset import

Figure 5 Next is importing data for preprocessing data. Data import is done by using the Pandas library and divided into 2 parts, i.e**X, representing features, and Y, representing labels (predicted results).**

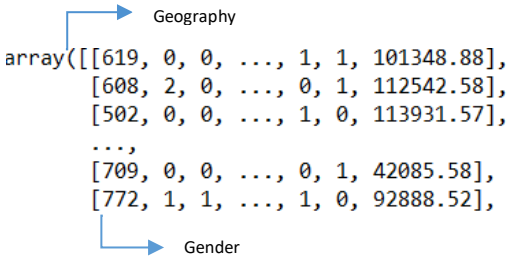

Fig 6. The result of the conversion becomes an integer

Figure 6: Some of the imported data is still in the form of categorical (string), i.e. geography and gender, which are required to be converted into integers to be processed by the model. This process is done using the LabelEncoder module (changing values to 0 and 1) from Scikit-learn.

X_test	float64	(2000, 11)	[[1.0000000e+00 0.0000000e+00 5.9700000e+02 ... 1.0000000e+00 1.0000 ...
X_train	float64	(8000, 11)	[[0.0000000e+00 1.0000000e+00 6.6700000e+02 ... 1.0000000e+00
y_test	int64	(2000,)	[0 1 0 ... 0 0 0]
y_train	int64	(8000,)	[0 0 0 ... 0 0 1]

Fig 7. Training data and testing data

Figure 7: The dataset is divided into 2, namely training data (with a total of 8000 records) and testing data (with a total of 2000 records). Data training is used to train neural networks in making predictions whether customers will survive or stop using bank services, while testing data is used to find out whether the model can make predictions from new data and measure how accurate predictions can be made by the model.

A. Designing a Model

In this case the model is designed with the aim of predicting whether customers will survive or stop using bank services, by providing input data consisting of 10 features, each of which features are credit score, geography (categorical), gender (categorical), age, tenure, balance, number of products, has a credit card (categorical), is an active member (categorical), estimated salary to produce output in the form of true or false.

Accredited "Rank 4" (Sprott 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

a) Determine and define the number of Layers and Nodes

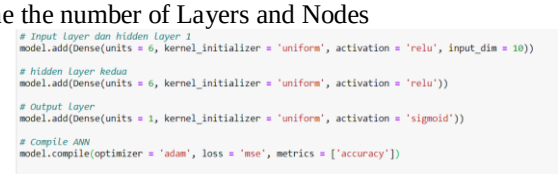

Fig 8. Layers and nodes

Model 1

Figure 8: The number of layers that will be used in the model, which is 4 layers consisting of input layer, 2 hidden layers, and the output layer with the number of each node in sequence is 10, 6, 6, 1. The number of hidden layers can be adjusted with how accurate the predicted results produced (which will be obtained in the process of testing the model), in this model the learning rate used is the default (0.01).

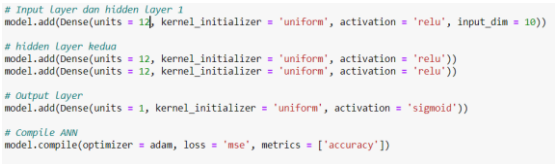

Fig 9. Layers and nodes

Model 2

Figure 9: The number of layers that will be used in the model, which is 5 layers consisting of input layer, 3 hidden layers, and the output layer with the number of each node in sequence is 10, 12, 12, 12, 1. The number of hidden layers can be adjusted to how accurate the predicted results produced (which will be obtained in the process of testing the model), in this model the learning rate used is the default (0.001).

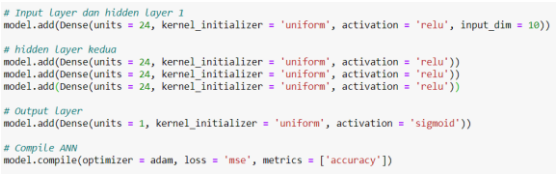

Fig 10. Layers and nodes

Model 3

Figure 10: The number of layers that will be used in the model, which is 5 layers consisting of input layer, 4 hidden layers, and the output layer with the number of each node in sequence is 10, 24, 24, 24, 24, 24, 1. Number hidden layer can be adjusted according to how accurate the predicted results produced (which will be obtained in the process of testing the model), in this model the learning rate used is the default (0.001).

Each hidden layer in a neural network has a weight variable on each connection between nodes. This variable is generated with a uniform function (ranging from 0 to 1), on each layer the value changes each time the training process (epoch) is carried out (adjusts to loss).That way, the data (input) and weight will be operated into a sequential matrix calculation system and changed with the activation function to generate output for each layer.

There are many activation functions that can be used on hard-tensorflow, one of the most commonly used is Rectified Linear Unit (ReLU) used in this model.

Loss and Accuracy are functions to measure how well a neural network predicts. Loss is obtained by running the Loss Functioon function. The Loss Function used is Mean Squared Error (MSE).

Optimizer is used to optimize the model's prediction results as accurately as possible. This function is run using Adam Optimizer, which is an optimizer commonly used in hard-tensorflow.

B. Train Model

Table 1

This table is the result of designing a model from 3 layer and node models that have been created

Accredited "Rank 4" (Sprott 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

	Accuracy of training	Validation accuracy
Model 1	83%	84%
Model 2	86%	86%
Model 3	87%	85%

Making 3 different models is done so that the model can study the correlation between feature data and labels (done with mathematical calculations), so that later it can be implemented on new data.

```
# Training model
model.fit(X_train, y_train, batch_size = 10, epochs = 100, shuffle=True)
```

By writing the model.fit function and writing some arguments, 10 feature data (X_train) to train the model, data label (Y_train), number of epochs (forward propagation and backpropagation) the model will be carried out during the training and shuffle process which functions to randomize the data in the training data (neural networks will learn better with random training data).

```
Epoch 1/100
8000/8000 [==============================] - 2s 218us/step - loss: 0.1565 - acc: 0.7960
Epoch 2/100
8000/8000 [==============================] - 1s 102us/step - loss: 0.1342 - acc: 0.7965
Epoch 99/100
8000/8000 [==============================] - 1s 109us/step - loss: 0.1232 - acc: 0.8359
Epoch 100/100
8000/8000 [==============================] - 1s 96us/step - loss: 0.1233 - acc: 0.8361
Out[4]: <keras.callbacks.History at 0x211845eef60>
```

Fig 11. Model training process

Figure 11: Model 1 training process (one sample taken from the three model designs) with an epoch of 100, with a loss of 0.1233 and an accuracy of 84%.

C. Test the model

The data used in this stage are X_test and Y_test from the dataset (as many as 2000 records) to test whether the model can predict from new data, and not only remember the data from the model training process.

	0	1
0	1547	48
1	269	136

```
In [17]: (1547+136)/2000
Out[17]: 0.8415
```

Fig 12. Accuracy obtained for testing data

Figure 12 Accuracy obtained in the testing process is the same as the accuracy obtained from the training process that is equal to 84%, which means the model can make predictions from new data and not only remember the data from the training process model.

D. Model Implementation

The next step is to implement the model that has been trained and tested. In this case the model implementation is done by deploying the model into web-based applications (web app), using the Flask Service contained in the Python programming language.Flask is a microframework that can be used in making web programming language Python, Flask is called a micro website because it does not require some additional tools / libraries. Flask does not have a database abstraction layer, validation form or some other third party component.

a) Deploy the Deep Learning Model

After the model can make good predictions, the model needs to be deployed so that it can be used by related companies to predict churn. This process is done using the Flask Service found in the Python programming language to deploy the model into web-based applications (web app).

The following is an illustration of the process that runs on the Bank Churn Prediction web-based application.

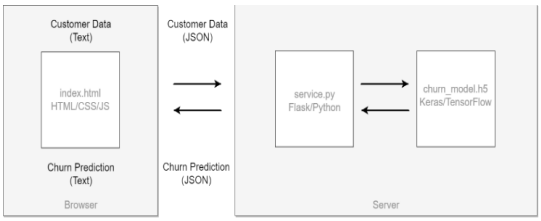

Fig 13. Illustration of processes running in an application

1) Users input data on the application page, the data inputted among them, credit score, geography (France, Spain, Germany), gender (Male, Female), age, tenure, balance, number of products, has cr cards (Yes, No), is active member (Yes, No), estimated salary, which is a feature data for making predictions.
2) After the predict button is clicked, the data will be sent through the localhost network using the HTTP POST method in the form of JSON format.
3) Flask service receives data, then runs the Hard predict function on data that has been received using the deep learning model, namely Churn_Model.h5.
4) After the model predicts output, the output data is sent back in JSON format via the localhost network.
5) Output is displayed on the application page as a prediction result.

The application is designed using HTML, CSS, and JavaScript as a frontend and uses Flask as a server to run on the localhost network.

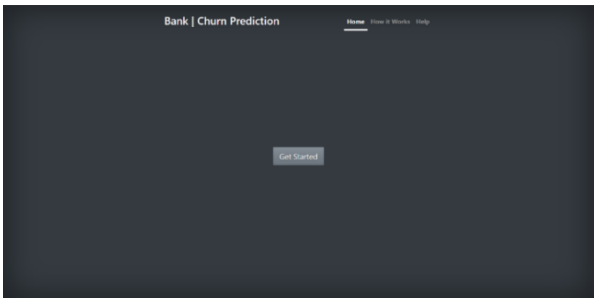

Fig 14. Home page

Figure 14 The Home page consists of several menus, each of which leads to a specific page, with a "Get Started" button, which leads to the main page of the application.

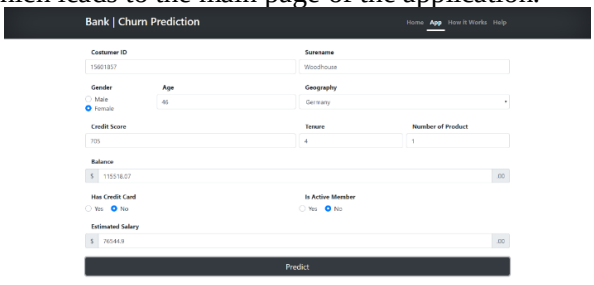

Fig 15. Main Page Application, where customer data is input

Figure 15 The main page of the application where users input customer data to predict whether the customer will continue to use (stay) or stop (leave). Data in sequence are Gender (Yes, No), Age, Geography (France, Spain, Germany), Credit Score, Tenure, Number of Products, Balance, Has Credit Card (Yes, No), Is Active Member (Yes, No) , and Estimated Salary.

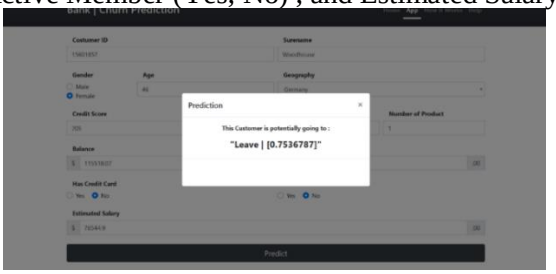

Fig 16. Prediction results for data input

Accredited "Rank 4" (Sprott 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

Figure 16 The results in the picture above are the results of predictions made by the model based on data that has been input, with an output in the form of a prediction that the customer will leave (stop using bank services).

4. Conclusion

With the application of a deep learning model based on Artificial Neural Network (ANN) to solve problems in the form of classification problems in the case of predicting the banking industry churn, the results obtained in the form of a web-based application (web app) that can predict the possibility of customers will stop or continue to use the services provided by banks . With customer data that has been inputted before, the model can make predictions with an accuracy of 84% and a loss of 0.1233. The model that was designed using the supervised machine learning method with a dataset in the form of bank customer data churn that has already existed, proves that this model can make churn predictions accurately enough to assist the company in making further decisions, so that losses can be avoided.

5. Reference

[1]. Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. October.
[2]. Arifin, M. (2015). Ig-Knn Untuk Prediksi Customer Churn Telekomunikasi. Simetris : Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer, 6(1), 1.
[3]. Bandi, R., & Amudhavel, J. (2018). Object recognition using Keras with backend tensor flow. International Journal of Engineering and Technology(UAE), 7(3.6 Special Issue 6), 229–233.
[4]. Buslim, N. (2019). Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial sebagai media promosi Online Bagi Masyarakat. Jurnal Teknik Informatika, 12(1), 79–96.
[5]. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Peter Campbell, J. (2020). Introduction to machine learning, neural networks, and deep learning. Translational Vision Science and Technology, 9(2), 1–12.
[6]. Goldsborough, P. (2016). A Tour of TensorFlow. October.
[7]. Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. Physics Procedia, 62, 18–22.
[8]. Jan, P., & Gotama, W. (2018). Pengenalan Pembelajaran Mesin dan Deep Learning. 2019, July, 1–199.
[9]. Kumar, V., & L., M. (2018). Deep Learning as a Frontier of Machine Learning: A Review. International Journal of Computer Applications, 182(1), 22–30.
[10]. Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. Risks, 7(1). https://doi.org/10.3390/risks7010029
[11]. Muhammad, I., & Yan, Z. (2015). Supervised Machine Learning Approaches: a Survey. ICTACT Journal on Soft Computing, 05(03), 946–952.
[12]. Pertiwi, R. W., & Widiyanto, I. (2015). Minat Churn Pelanggan Indosat Di Indonesia. 4, 1–13.
[13]. Rachmat, A. (2019). Survei Penerapan Model Machine Learning Dalam Bidang Keamanan Informasi. Jurnal Sistem Cerdas, 2(1), 47–60.
[14]. Ritonga, A. S., & Atmojo, S. (2018). Pengembangan Model Jaringan Syaraf Tiruan untuk Memprediksi Jumlah Mahasiswa Baru di PTS Surabaya (Studi Kasus Universitas Wijaya Putra). Jurnal Ilmiah Teknologi Informasi Asia, 12(1), 15.
[15]. Vogt, M. (2019). An Overview of Deep Learning and Its Applications. January, 178–202. https://doi.org/10.1007/978-3-658-23751-6_17