**CHAPTER THREE**

**MATERIALS AND METHODS**

**3.1 Choosing the Best Environment**

Knowing the data analytics tools and the intricacy of the Volve dataset, the Python programming language was selected to develop models and conduct computations. Python is often regarded as the greatest scripting language for data mining and analysis. Furthermore, it has a large community, so any difficulties encountered may be readily addressed owing to the knowledge given on specialized forums. This ecosystem includes a plethora of strong libraries spanning from fundamental statistics to complicated machine learning techniques [21]. All libraries shine in terms of performance, productivity, and collaboration, making the entire process of data handling and visualization relatively simple as compared to other languages.

**3.2. Data Acquisition**

Data is a major factor in artificial neural network's forecasting ability. In other words, ANN processes a large volume of data that spans a wide range of operational circumstances. If poor, noisy, and insignificant amounts of data are expected, ANN is clearly not the best option. For pattern recognition, prediction, system identification, and control, ANNs have emerged as the technology of choice because of their capacity to handle complicated, non-linear connections between parameters.

The initial hurdle was familiarizing oneself with the available data due to the enormous quantity of the collection.

The oil well production data includes the date of production, codes and names as determined by the Norwegian Petroleum Directorate, including the well field code, name, and facility. Following these are the following: Pressure Data (Downhole Pressure & Temperature, Average Tubing & Annular Pressure Details), Flow Data (Volumetric Data for Produced Gas, Oil, Water, and Injected Water), Flow Kind (e.g. Spherical), and Other Data (Onstream Hours, Type of Well, Choke Details (e.g. Choke Size, DP Ratio [Pressure drop/downstream pressure]),were gathered from an abandoned field since the validity of the data is crucial to the accuracy of the machine learning process.

**3.3 Selection of Best Input and Target Parameters**

It is generally agreed that in neural networks, having enough data available is one thing, but knowing the best parameters to use for the best outcome is another, as having more input parameters than necessary will result in a large network size and a corresponding decrease in learning speed and efficiency (Goda et al., 2005).

As a result of the critical analysis and independence examination, ON STREAM HRS, AVG DOWNHOLE PRESSURE, AVG DP TUBING, AVG WHP P, AVG WHT P were used as primary input parameters, while BORE OIL VOL was the target output, as knowing the full influence of the parameters is the primary point of concern in ANN model development.

**3.3 Data Pre-Processing**

Normalization method before presenting the input data to the network is mostly a decent drill. Mixing variables with large amounts and small amounts blur the  
learning algorithm on the prominence of an individual variable and may drive it to finally discard the variable with the lesser amount (Tymvios et al., 2008). To ensure that the selected data for modeling distinguish distinctive operational varieties and off data set will be separated from the list. The data set will be grouped in two i.e. training, and testing data sets.

Consequently, the considered parameters were normalized, using the min-max method, to speed up learning and achieve convergence and in order for the neural network to be more efficient and, in overall, yield better predictions.

Hence, the available data were normalized into the range of 0 to 1 by using Equation 3.1 shown here:

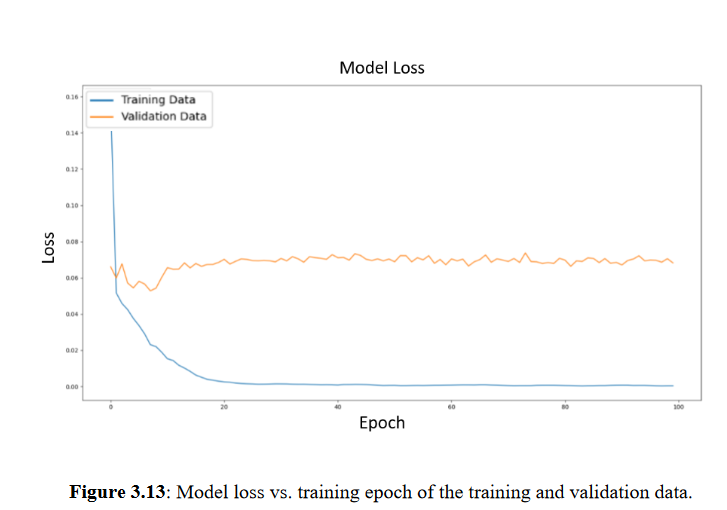
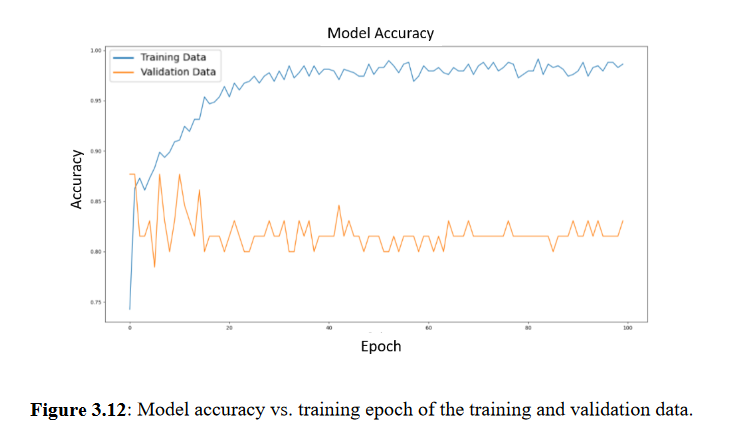
Equation 3.1

Where Xn = normalized value, Xmin = minimum of original values, Xmax = maximum of original values, and X = original value

**3.4. Training of an Artificial Neural Network**

It is an optimization procedure where an error function is decreased by regulating  
ANN weights. In the training period, a set of recognized input-output configurations are continually offered to the network in order to demonstrate the network. The weights are constantly adjusted in order to decrease the error. The system learns how to handle different situations so that the network model can appropriately handle the similar scenarios in the future. Thus error is deduced from comparing output of an ANN with the actual output provided by the user. The training is to be conducted until error is reduced and better performance of ANN is achieved. A both stable and convergent ANN is desired from a training phase.

To train the network, Python was used along with the TensorFlow and Keras libraries. The normalized inputs and target were selected and fed to the system. The network was trained over 100 epochs with a learning rate of 0.001 with the Adam optimizer. The Adam optimizer combines the best traits of other optimizers that use root mean square methods to improve their ability to work with noisy data. The default parameters of the Adam optimizer work well enough to train networks effectively [81]. Figures 3.12 and 3.13 highlight the improvements in model accuracy and model loss of the network throughout training.



**3.5 Design of Network Architecture**

The number of neurons in the input and output layers quite simply determine the number of input and output parameters. ANN model typically consists of three layers: input, hidden, and output layers. Each of these layers consists of neurons (sometimes called nodes). The input layer is the first layer which has neurons that represent the inputs or parameters of a certain problem. The second layer is the hidden layer where the computational process is initiated. The output layer represents the labels or outputs of the problem, and consists of certain number of neurons based on the nature of the problem, being one—probability of pipe sticking—in our situation. In the work proper, well-thought-out choice of number of hidden layer(s) of just 1 with 5 neurons was made.