CHAPTER 3
RESEARCH METHODOLOGY

3.1.   Data Collection

The dataset used in this project are from the experiments of Agricultural Technology, Soegijapranata Catholic University students, Alice Septiana Dewi and Irmadella Rana Nathania, in their research. This dataset has 6 types of microplastics with references to absorption bands from several literature and the absorption bands themselves. The types of microplastics are PA/Nylon, PP, PS, PVC, PE/LDPE, and PET/PETE.

The dataset has their references for each microplastics. They gathered it from several references, one of them is M. R. Jung et al. [6] The paper has summarized many microplastics and its absorption band polymer.

3.2.   Data Preprocessing

There are several steps for preprocessing the dataset. First, split Data Using K-fold. Dataset from Faculty of Agricultural Technology are splitted using K-fold. K-fold is used because of the limited data for the dataset. Second, calculate the mean of gaussian mixture. References from literature become predefined to gaussian mixture. Fit the gaussian mixture model with the data train that has been splitted with K-fold. Lastly, identify the polymers. After the gaussian mixture model is created, the absorption band data are converted into a polymer with probability. The maximum probability of the polymer is selected into an array.

For experiment, data augmentation is used to compare the performance of the model. Crossover become the base idea for the augmentation. The crossover of 2 data from the same microplastic will generate several data.

3.3.   Experiment

Experiments were conducted with 2 different methods, Manual Matching and Gaussian Mixture Naïve Bayes matching.

3.3.1. Manual Matching

Each absorption band from the FTIR result is matched to the reference of polymer absorption bands. The user must estimate the value of the absorption band into polymer reference.

3.3.2. Gaussian Mixture and Naive Bayes Matching

All data are converted into a csv file separated from the references and the dataset. The data were processed in gaussian mixture model to identify the polymer and gaussian naive bayes to classify the microplastics types (PA/Nylon, PP, PS, PVC, PE/LDPE, PET/PETE).

3.4. Evaluation

For the evaluation, the author used K-fold to check the performance of the Gaussian Naive Bayes Model. The dataset was split into 6 folds and iterated the process The performance was measured with Classification Report by Sklearn.metrics. The report contains precision, recall, f1-score, support, and accuracy. As an addition, the performance was compared to another algorithm.

3.5. Discussion

The results of this paper will be the range of absorption band scope based on gaussian mixture model. This will be used as a reference to know the standard deviation of absorption bands polymer around the other reference. Furthermore, it can be used for any identification of polymer absorption band as long as has the reference absorption band. This also can be used to classify the other microplastics with adjustment on the reference.