

```

% KLASIFIKASI TEXT MENGGUNAKAN DEEP LEARNING

filename = "weatherReports.csv";
data = readtable(filename,'TextType','string');
head(data);

idxEmpty = strlength(data.event_narrative) == 0;
data(idxEmpty,:) = [];

data.event_type = categorical(data.event_type);

f = figure;

f.Position(3) = 1.5*f.Position(3);
h = histogram(data.event_type);
xlabel("Class")
ylabel("Frequency")
title("Class Distribution")

classCounts = h.BinCounts;
classNames = h.Categories;

idxLowCounts = classCounts < 10;
infrequentClasses = classNames(idxLowCounts)

%% infrequentClasses = 1x8 cell array {'Freezing Fog'}      {'Hurricane'}
{'Lakeshore Flood'}      {'Marine Dense Fog'}      {'Marine Strong Wind'}
{'Marine Tropical Depression'}      {'Seiche'}      {'Sneakerwave'}

idxInfrequent = ismember(data.event_type,infrequentClasses);
data(idxInfrequent,:) = [];
data.event_type = removecats(data.event_type);

cvp = cvpartition(data.event_type,'Holdout',0.1);
dataTrain = data(training(cvp),:);
dataTest = data(test(cvp),:);

textDataTrain = dataTrain.event_narrative;
textDataTest = dataTest.event_narrative;
YTrain = dataTrain.event_type;
YTest = dataTest.event_type;

figure;
wordcloud(textDataTrain);
title("Training Data")

textDataTrain = erasePunctuation(textDataTrain);
textDataTrain = lower(textDataTrain);
documentsTrain = tokenizedDocument(textDataTrain);

```

```
%% documentsTrain(1:5)

embeddingDimension = 100;
embeddingEpochs = 50;

emb = trainWordEmbedding(documentsTrain, ...
    'Dimension',embeddingDimension, ...
    'NumEpochs',embeddingEpochs, ...
    'Verbose',0)

documentLengths = doclength(documentsTrain);
figure
histogram(documentLengths)
title("Document Lengths")
xlabel("Length")
ylabel("Number of Documents")
```