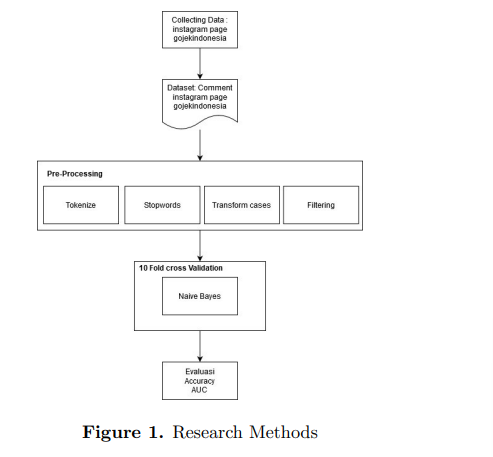
Sentiment analysis on passengers feed back & creating Dashboard

Abstract: Sentiment analysis is a computational study of opinions and emotions expressed textually. Sentiment analysis will group text in sentences or documents to find out the opinions expressed in the sentence or document, whether negative or positive. This sentiment analysis research was conducted on the online taxi transportation service. Gojek uses a lot of social media to communicate with its customers, one of the social media used is Instagram. This research takes 1000 comments from the Instagram of the Gojek page which is used to see the public opinion of the online Gojek transportation services. Comments from the page are processed by doing text preprocessing and then classified using the Naive Bayes Classifier (NBC) method to obtain the value of the public value for online transportation services. The results of this study using the Naive Bayes method resulted in an accuracy value of 81.00%, which means that from all the comments on the Instagram page, the subject of the NBC method could be accurately classified by 81.00% whether the comments were negative or positive comments.

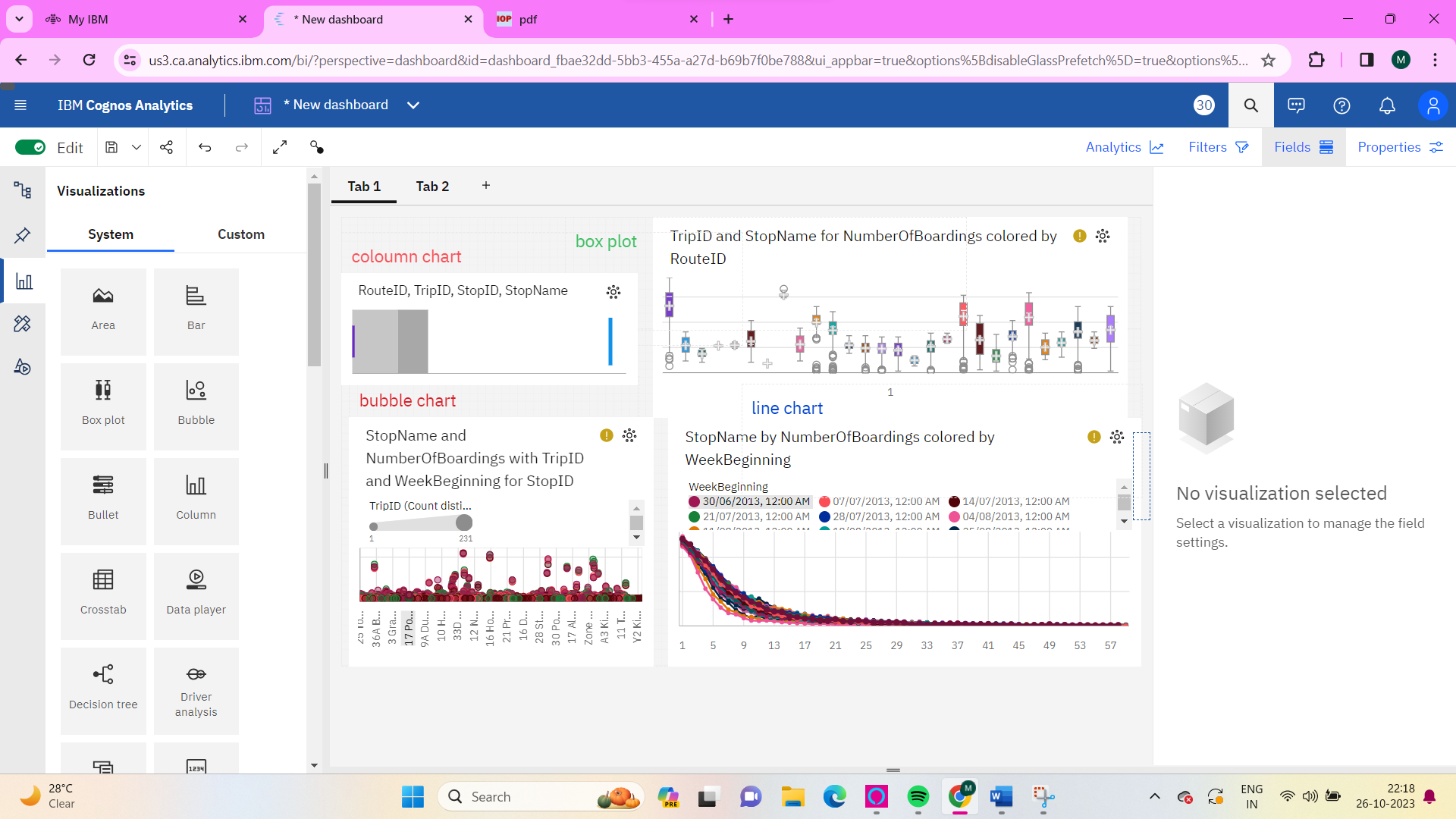
Research Methods : Sentiment Analysis Process This research is divided several steps. First, data is collected automatically from the Gojekindonesia Instagram page. Raw data is prepared for subsequent manuals. Furthermore, pre-processing consists of tokenization, stopwords, transform case, and filtering. After the data is ready in the next process is a weighting process based on the number of words that appearin the document, the feature used is the Frequency Document Reverse Frequency (TF-IDF) term. Next is the classification of comment data using the Na¨ıve Bayes Classifier. Evaluation is done by comparing the predicted labels with the labels calculated by the system. Review of the research that will be carried out in Figure 1.



Validation:

Validation in the process of data Cross Validation is divided into two parts. the first part is a subset of training and the second part is a subset of tests that will be used to test performance. By using K-fold cross-validation, the data will be automatically partitioned, the original dataset becoming subsample K. One subsample is stored as validation data to test the classification model and the remaining subsample is used for training data in Subsample K. The validation process is repeated continuously for K with each k sample used once as test data. Then to produce an estimate of performance of results K from the fold then averaged. To get the best model, the dataset distribution process is done using 10 times Cross-Validations.

The resulting rule will be used as the basis for predicting the value to be performed. Previously the rule had to be evaluated and validated so that it was known how accurate the prediction results would be. At this stage the evaluation process uses the confusion matrix method. The confusion matrix model will form a matrix consisting of accuracy, recall, and precision.



Area chart

