

- Problem:  
Babe app is the most popular news and video application in Indonesia. There are thousands of new comments on Babe application everyday. But many comments are Sara comments which have religious discrimination and blasphemy content.
- Task:  
Your task is to develop a computer program which can detect Sara comments automatically.
- Data
  - Indonesian Dictionary: The list of the most frequent words in Indonesia (202115 words).
  - Indonesian stop words: List of stop words in Indonesia (7224 words).
  - Training set: there are 60392 normal comments and 14908 Sara comments.
  - Test set: there are 6830 normal comments and 2816 Sara comments.
- Guidances
  - You can use these programming languages: Python, Scala, Java or C++/C.
  - You can use any machine learning models, rule-based algorithms, or ensemble of models.
  - You should write a code to compute some evaluation metrics: accuracy, precision, recall, and F1-score.
  - You have to write a document to explain your solution and how to run model training and testing.