

[DOI: 10.20472/IAC.2018.935.054](https://doi.org/10.20472/IAC.2018.935.054)

**FARAH ZAWAIDEH**

Irbid National University, Jordan

**RAED SAHAWNEH**

Irbid National University, Jordan

## **IMPROVE NAÏVE BAYESIAN CLASSIFIER BY USING GENETIC ALGORITHM FOR ARABIC DOCUMENT**

### **Abstract:**

Automatic text categorization (TC) has become one of the most interesting fields for researchers in data mining, information retrieval, web text mining, as well as natural language processing paradigms due to the vast number of new documents being retrieved for various information retrieval systems. This paper proposes a new TC technique, which classifies Arabic language text documents using the naïve Bayesian classifier attached to a genetic algorithm, model; this algorithm classifies documents by generating a random sample of chromosomes that represent documents in the corpus.

The developed model aims to enhance the work of naïve Bayesian classifier through applying the genetic algorithm model. Experiment results show that the precision and recall are increased when testing higher number of documents; the precision was ranged from 0.8 to 0.97 for different testing environment; the number of genes that is placed in every chromosome is also tested and experiments show that the best value for the number of genes is 50 genes

### **Keywords:**

Data mining, Text classification, Genetic algorithm, Naïve Bayesian Classifier, N-gram processing

**JEL Classification:** C80