# تطبيق خوارزمية تحسين الحيتان والخوارزمية الجينية لكشف المحتوى الغير مرغوب فيه في تويتر

**الطالبة: فاطمة محمد منصور القحطاني**

**المشرفة: غادة أحمد عامودي**

## المستخلص

على مدى السنوات العشر الماضية ، أصبحت شبكات التواصل الاجتماعي عبر الإنترنت أكثر شيوعًا. يزيد استخدام مواقع التواصل الاجتماعي المختلفة مثل ( Facebook و Twitter و Instagram و Snapchat و Pinterest و LinkedIn ) من القدرة على نشر المعلومات من خلال شبكات التواصل الاجتماعي. كما أنه يسهل انتشار المحتوى العشوائي ويهدد معلومات المستخدم وخصوصيته. تم إجراء العديد من الدراسات البحثية للكشف عن المحتوى المزعج في شبكات التواصل الاجتماعي . في هذه الدراسات ، يعتمد الباحثون على عمليات التصنيف التي تستخدم عددًا كبيرًا من الميزات ، مما يتطلب وقتًا أطول للتنفيذ. يقترح هذا البحث أطروحة جديدة للكشف عن التغريدات المزعجة في تويتر باستخدام مجموعتي بيانات لغويتين مختلفتين (الإنجليزية والعربية). يستخدم النموذج المقترح طريقة تصنيف محسنة لاكتشاف المحتوى المزعج في التغريدات. لعملية التصنيف ، يتم استخدام ثلاث خوارزميات ( Naïve Bayes, Logistic Regression ,Stochastic Gradient Descent). ثم يتم التحسين في تجربتين منفصلتين. في التجربة الأولى ، تم تطبيق خوارزمية Whale Optimization Algorithm. في الثانية ، يتم تطبيق الخوارزمية الجينية Genetic Algorithm. أسفرت تجارب البحث عن النتائج التالية: بالنسبة لمجموعة البيانات الإنجليزية ، كانت خوارزمية Naïve Bayes هي أفضل مصنف لنتائجها العالية قبل التحسين و بعد التحسين ، وتحسنت دقتها من ٩٣,١٪ إلى ٩٥,٣٪ بعد التحسين. بالإضافة إلى ذلك ، انخفض عدد الميزات المطلوبة من ٢٠٠٠٠ إلى ٣٠٠٠. بالنسبة لمجموعة البيانات العربية ، كانت خوارزمية Logistic Regression أفضل مصنف قبل وبعد التحسين ، تحسنت دقتها من ٨٩,٥٪ إلى ٩١,١٪. على غرار مجموعة البيانات الإنجليزية ، انخفض عدد الميزات المطلوبة من ٦٦٨٩ إلى ٢٤٠٠. وأظهرت النتائج أنه بالنسبة لمجموعتي البيانات ، فإن Whale Optimization Algorithm عزز دقة عملية التصنيف ، وخفض عدد الميزات المطلوبة. وبالتالي ، يمكننا أن نستنتج أن تحسين عملية التصنيف باستخدام خوارزميةWhale Optimization Algorithm يحسن نموذج التصنيف ، مما يؤدي إلى توقع أسرع في الوقت الفعلي ووقت تنفيذ أقصر. بينما استخدام Genetic Algorithm لم يحسن نتيجة نموذج التصنيف للتجارب التي أجريت في هذا البحث.

# Applying Whale Optimization and Genetic Algorithms for Spam Detection in Twitter

**By**

**Fatimah Mohammed Mansour Alqahtani**

**Supervised by**
**Dr. Ghada Amoudi**

## المستخلص

Over the past 10 years, Online Social Networks (OSNs) have become more popular. The use of various social media sites, such as Facebook, Twitter, Instagram, Snapchat, Pinterest, and LinkedIn, increases the ability to disseminate information through OSNs. It also facilitates the spread of spam and threatens user information and privacy. Many research studies have been conducted to detect spam in OSNs. In these studies, the researchers depend on classification processes that use a large number of features, requiring longer execution time. This thesis research proposes a new model to detect spam in Twitter using two different linguistic datasets (English and Arabic). The proposed model uses an optimized classification method to detect spam in tweets in the two different linguistic datasets. For classification process, three algorithms are used Naïve Bayes, Logistic Regression and Stochastic Gradient Descent. Then the optimization is done in two separate experiments. In the first experiment, the Whale Optimization Algorithm (WOA) is applied. In the second one, the Genetic Algorithm (GA) is applied. The research experiments yielded the following results: for the English dataset, the Naïve Bayes algorithm was the best classifier for its high results before optimization, and it had the highest results after optimization, its accuracy improved from 93.1% to 95.3% after optimization. Additionally, the number of required features decreased

from 20,000 to 3000. For the Arabic dataset, the Logistic Regression algorithm was the best classifier before and after optimization, its accuracy improved from 89.5% to 91.1%. Similar to English dataset, the number of required features decreased from 6689 to 2400. The results showed that for both datasets, the WOA enhanced the accuracy of the classification process, and it reduced the number of the required features. Thus, we can conclude that optimizing the classification process using the WOA algorithm improves the classification model, leading to faster real-time prediction and a shorter execution time. While using the GA did not improve the result of the classification model for the experiments done in this research.