

# التصنيف التلقائي لتغريدات الجهات المختلفة لجامعة الملك عبد العزيز

الطالبة

ولاء صالح الحبشي

المشرف

د. نايف راضي الجهني

**الملخص:** وسائل التواصل الاجتماعي هي المكان الرئيسي للتفاعل الاجتماعي اليوم. أحد هذه الوسائط هو Twitter الذي أصبحت خصائصه واستخداماته محورًا لمختلف المجالات مثل الرعاية الصحية والاتصالات والتعليم. بما أن التعليم هو المجال الرئيسي لتنمية المجتمعات، فقد اهتمت المؤسسات التعليمية بقياس تفاعلها مع طلابها على تويتر لتحسين أداء الخدمات التعليمية من خلال تصنيف التغريدات. يعد تصنيف التغريدات عبر خوارزميات التعلم الآلي وسيلة قابلة للتطبيق لتقييم فعالية حملات وسائل التواصل الاجتماعي والاتصالات. تهدف الدراسة الحالية إلى فحص مجال التنقيب على مواقع التواصل الاجتماعي من خلال تصنيف التغريدات باللغة العربية لمختلف قطاعات جامعة الملك عبد العزيز، وهي مؤسسة تعليمية في المملكة العربية السعودية، من خلال بناء إطار لتصنيف التغريدات إلى فئات محددة مثل القبول. والامتحانات والمكتبات. يفحص هذا الإطار قوة السمات النمطية باستخدام ثلاث مصنفات مختلفة للمدونات الصغيرة العربية. السمات النمطية المطبقة على النص في هذا الإطار هي الشخصية والمعجمية والنحوية. تم استخراج هذه الميزات باستخدام أداة MADAMIRA واستخدمت لبناء نموذج تصنيف قائم على الميزات. في هذا النموذج، يتم دمج الميزات مع مجموعة أحجام مختلفة من البيانات باستخدام ثلاث تقنيات تصنيف وهي (support vector machine (SVM، k-nearest neighbour و random forests.

تم استخدام تقنية K-folds لتقييم النماذج، وتم إجراء مقاييس الدقة والدقة والتذكر والقياس f فيما يتعلق بكل نموذج تعلم آلي. قارنت الدراسة النماذج الثلاثة للتأكد من النماذج التي تقدم أفضل أداء من حيث الدقة. تظهر النتائج أن SVM هو المصنف الأفضل أداءً. بينما أثبتت الغابة العشوائية أيضًا أنها مصنف قوي، إلا أنها لم تحقق أداءً جيدًا مثل SVM.

# **AUTOMATIC CATEGORIZATION OF KING ABDUL-AZIZ UNIVERSITY SECTORS' TWEETS**

**By  
Walaa Saleh Alhebshi**

**Supervised by  
Dr Naif Radi Aljohani**

**Abstract:** Social media are the key venue of social interaction today. One such medium is Twitter whose characteristics and use has become the focus of various fields such healthcare, communications, and education. As education is the main field for the development of societies, educational institutions have paid attention to measuring their interaction with their students on Twitter to improve the performance of educational services by classifying tweets. Classifying tweets via machine learning algorithms is a viable means of evaluating the effectiveness of social media campaigns and communications.

The present study is aimed at examining the field of social network mining through classifying the tweets in Arabic of various sectors of King Abdul-Aziz University, an educational institution in Saudi Arabia, through building a framework to classify the tweets into specific categories such as admission, exams, and libraries. This framework examines the power of stylometric features using three different classifiers of Arabic microblog. The stylometric features applied to the text in this framework are character, lexical, and syntactic. These features were extracted using the MADAMIRA tool and were used to build a features-based classification model. In this model, the features are combined with a different size set of data using three classification techniques, which are support vector machine (SVM), k-nearest neighbour, and random forests.

The K-folds technique was used to evaluate the models, and the measures of precision, accuracy, recall, and f-measure were made regarding every machine-learning model. The study compared the three models to ascertain which models would offer the best performance in terms of accuracy. The findings show that SVM is the best performing classifier. While the random forest also proved to be a strong classifier, it did not perform as well as the SVM.