

# نموذج تصفية مشخصة مختلط للمحتويات غير المرغوبة في وسائل التواصل الاجتماعي

اسم معد الرسالة  
مشاعل بنت محمد السلمي

بحث مقدم لنيل درجة الدكتوراه في العلوم  
( علوم الحاسبات )

اسم المشرف على الرسالة  
أ.د. أروى بنت يوسف الأعمى

كلية الحاسبات وتقنية المعلومات  
جامعة الملك عبدالعزيز  
جدة . المملكة العربية السعودية  
العام الجامعي ٢٠٢٠ / ٢٠٢١

## المستخلص

إن المحتوى الكبير الذي ينشئه المستخدمين والناجم عن الاستخدام الشائع لخدمات التواصل الاجتماعي يُعرض المستخدمين لأنواع مختلفة من المحتوى الذي قد يكون ضاراً أو غير مرغوب فيه. لا يمكن تعميم حلول حماية خصوصية المستخدم من مثل هذا المحتوى الغير مرغوب فيه بسبب التصورات المختلفة لما يُعتبر غير مرغوب فيه لكل فرد. أيضاً هناك سبب آخر لعدم قابلية التعميم وذلك يرجع إلى الاختلافات في القيم الثقافية بين المجتمعات، والتي يتم مشاركتها بين المستخدمين من نفس المجتمع.

انطلاقاً من القيود و التحديات في أساليب التصفية وحماية الخصوصية الحالية، يهدف هذا العمل إلى توفير نموذج تصفية شخصي مختلط يدمج التحليل السياقي، السلوكي، العاطفي، والثقافي للكشف عن المحتوى الغير مرغوب فيه للأفراد في خدمات التواصل الاجتماعي. سيدمج النموذج المقترح كلاً من تقنيات تعلم الآلة وتنقيب البيانات لتحقيق هدفه.

في هذا العمل أجرت الباحثة ثلاث دراسات تجريبية. كانت دراسة المستخدم الأولى بمثابة دراسة تجريبية أكدت على الحاجة إلى حل مشكلة المحتوى الغير مرغوب فيه. شملت الدراسة الثانية ٣٩٤ مشاركاً وكان هدفها تحديد فئات المحتوى الغير مرغوب فيه وايجاد العلاقة بين اختيارات المستخدمين للمحتوى الغير مرغوب وسماتهم الشخصية من خلال تحليل الارتباط والتبعية. استخدمت الدراسة التجريبية النهائية منصة تويتر لاستخراج ومعالجة الجداول الزمنية لعينة من المستخدمين الحقيقيين لتقييم النموذج المقترح ببيانات حقيقية.

للتحقق من صحة النتائج النهائية لنموذج التصنيفية الشخصية المختلط، قارنت الباحثة النتائج المستخرجة من النموذج مع ملاحظات المستخدم الصريحة وشبه الصريحة باستخدام الاختبارات الإحصائية. كشفت النتائج عن اتفاق بينهما. أظهرت النتائج أن العمر هو العامل الشخصي الأكثر ارتباطاً باختيار المستخدمين للمحتوى الغير مرغوب. ايضاً اظهرت النتائج ان سلوك التعليق على الرسائل هو السلوك الاكثر تأثيراً على اختيارات المستخدم للمحتوى الغير مرغوب.

يتكون نموذج التصنيفية الشخصية المختلط من ثلاث نماذج تصنيف: نموذج التصنيف المستند إلى تحليل المشاعر، نموذج التصنيف المستند إلى الثقافة، ونموذج التصنيف العربي. تم اعتماد مجموعة متنوعة من مصنفات التعلم الآلي لتحقيق أفضل نتائج أداء وتم تقييمها باستخدام مقاييس الأداء والتصنيف.

أخيراً، يمكن أن يكون للنموذج المقترح دور فعال في تطوير حلول مخصصة لمشكلة المحتوى غير المرغوب، مما يزيد من جودة التجربة لمستخدمي وسائل التواصل الاجتماعي.

# **A Hybrid Personalized Filtering Model for Unwanted Content in Online Social Network Services**

By  
Masha'el Mohammad Alsulami

Supervised by  
Professor Arwa Yousef Al-Aama

This thesis has been submitted in partial  
fulfillment of the requirements for the degree of  
PhD of Science in Computer Science

King Abdulaziz University, Jeddah  
College of Computing and Information Technology  
Department of Computer Science  
1442 H, 2021 G

## **ABSTRACT**

The high volume of user-generated content caused by the popular use of Online Social Network (OSN) services exposes users to different kinds of content that can be harmful

or unwanted. Solutions to protect user privacy from such unwanted content cannot be generalized due to different perceptions of what is considered as unwanted for each individual and due to the differences in cultural values among societies, which are normally shared among users from the same society.

Current solutions are either generalized the definition of unwanted content or suffer from the lack of adaptability in the filtering process. Driven by the limitations and challenges in current privacy protection and filtering methods, this work aims at providing a Hybrid Personalized Filtering Model that integrates behavioral, sentimental, and culture-based contextual analysis to detect unwanted content for individuals in OSN services. The hybrid model is proposed to benefit from machine learning, rule-based, and dictionary-based techniques to achieve its goal.

In this work, three experimental studies were conducted. The first user study acted as a pilot study that emphasized the demand to solve the problem of unwanted content. The second study involved 394 participants and its aim was to define the classes of unwanted content and to investigate the relationship between the users' choices of unwanted content and demographic features through correlation and dependency analysis. The final experimental study used Twitter API to extract and process the timelines of a sample of real users to evaluate the proposed model with real data.

To validate the final results of the proposed Hybrid Personalized Filtering Model, the outcomes derived from the proposed model were compared with user explicit and semi-explicit feedback using statistical tests. Results revealed an accurate agreement between them. Results also showed that commenting behavior is an effective positive indicator in detecting unwanted content for an individual. Moreover, results showed that age is the most related demographic factor to the users' choice of unwanted content.

The proposed model consists of three classification models: A Sentiment Analysis-Based Classification Model, and Topic-Based Classification Model that uses a supervised machine learning approach, and A Culture-Based Classification Model that uses a dictionary-based approach using the Lucene framework. A variety of Machine Learning classifiers were adopted to achieve the best performance results.

Finally, the proposed model can be used in developing personalized solutions to the problem of unwanted content, to increase the Quality of Experience (QoE) for users of OSN services.