

الاستفادة من تحليل بيانات وسائل التواصل الاجتماعي في تحسين الرعاية الصحية

شعيع دليم صندل العتيبي

المستخلص

بناء المدن الذكية يعتمد على قدرتنا على التفاعل مع بيئاتنا، وتحليلها، واتخاذ القرارات، كل ذلك في الوقت المناسب.

الرعاية الصحية هي أحد أهم المجالات التي تتطلب القدرة التحويلية لهذا الذكاء. وسائل التواصل الاجتماعي توفر بيئة مشاركة نشطة ومستمرة بين كافة أفراد المجتمع وصناع القرار في مجال الرعاية الصحية، مما يؤدي إلى تحسين الصحة العامة.

النظام المقترح في هذا البحث (صحة) هو نظام قائم أولاً على الاستفادة من تحليل المحتوى العربي لأحد منصات وسائل التواصل الاجتماعي (تويتر تحديداً). النظام يهدف إلى تحسين الرعاية الصحية في المملكة العربية السعودية،

تم بناء هذا النظام بالاعتماد على خوارزميات تعليم الآلة (naïve Bayes) وتقنية (Logistic Regression) المخصصة لتحليل البيانات الضخمة. وقد أظهر تطبيق النظام على مجموعة ضخمة من التغريدات (١٨,٩ مليون) نتائج مقارنة للإحصائيات المنشورة بعد مقارنتها بها. فقد تم اكتشاف قائمة أكثر خمسة أمراض شيوعاً بين أفراد المجتمع السعودي وفقاً لتغريداتهم

وهي أمراض الجلد وأمراض القلب وارتفاع ضغط الدم والسرطان والسكري. وأظهرت النتائج أيضا حاجة مدينتي الرياض وجدة إلى بذل المزيد من الجهد في مجال التوعية الصحية فيما يتعلق بهذه الأمراض الأكثر شيوعا.

الاستفادة من تحليل بيانات وسائل التواصل الاجتماعي في تحسين الرعاية الصحية

شعيع دليم صندل العتيبي

بحث مقدم لنيل درجة الدكتوراة في العلوم
(علوم الحاسبات)

الأستاذ الدكتور راشد محمود
الدكتور إياد كاتب

كلية الحاسبات وتقنية المعلومات
جامعة الملك عبد العزيز
جدة - المملكة العربية السعودية
جماد الآخرة ١٤٤١هـ - فبراير 2020 م

Leveraging Social Media Analytics for Enhanced Healthcare

Shoayee Dlain Alotaibi

ABSTRACT

Smartness, which underpins smart cities and societies, is defined by our ability to engage with our environments, analyze them, and make decisions, all in a timely manner. Healthcare is the prime candidate needing the transformative capability of this smartness. Social media could enable a ubiquitous and continuous engagement between healthcare stakeholders, leading to better public health. Current works are limited in their scope, functionality, and scalability.

This thesis proposes Sehaa, a big data analytics tool for healthcare in the Kingdom of Saudi Arabia (KSA) using Twitter data in Arabic. Sehaa uses Naive Bayes, Logistic Regression, and multiple feature extraction methods to detect various diseases in the KSA. Sehaa found that the top five diseases in Saudi Arabia in terms of the actual afflicted cases are dermal diseases, heart diseases, hypertension, cancer, and diabetes. Riyadh and Jeddah need to do more in creating awareness about the top diseases. Taif is the healthiest city in the KSA in terms of the detected diseases and awareness activities.

Moreover, we detected list of associated diseases, awareness events and measured the influence of top five health influencers in Saudi Arabia. Sehaa is developed over Apache Spark allowing true scalability. The dataset used comprises 18.9 million tweets collected from November 2018 to September 2019. The results are evaluated using well-known numerical criteria (Accuracy and F1-Score) and are validated against externally available statistics.