# تحليل المشاعر على مستوى الخدمات واستخراج أسم المكان لتغريدات اللغة العربية باستخدام التعلم العميق

## حالة دراسة: تغريدات شركات الاتصالات السعودية

**الطالبة / نوره بنت فهد بن علي الشمري**

**تحت إشراف**
**د.أمل عبدالله المنصور**

## المستخلص

في السنوات القليلة الماضية، أصبحت منصة تويترالأكثر شعبية للأفراد لمشاركة وجهات نظرهم وتجاربهم تجاه الخدمات والمنتجات المختلفة. لذلك، يجذب الكثير من الباحثين لاستخدامه كمرجع لاستخراج الرأي وتحليل المشاعر. تستخدم معظم الدراسات البحثية السابقة في هذا المجال النهج التقليدي القائم على التعلم الآلي والنهج القائم على المعجم لتصنيف الحالات العاطفية للتغريدات باللغة الإنجليزية. تم القيام بعمل بحثي محدود لتحديد توجهات الرأي للتغريدات بلغات أخرى مثل العربية. بالإضافة إلى ذلك، حقق نهج التعلم العميق مؤخرًا نتائج ملحوظة مقارنة بخوارزميات التعلم الآلي التقليدية في تحليل كمية هائلة من البيانات كما هو الحال مع بيانات الشبكات الاجتماعية. في هذه الرسالة، نسعى لتحديد ما إذا كان يمكن اعتماد نهج التعلم العميق لتعزيز أداء تحليل المشاعر للتغريدات العربية. لذلك، قمنا بتقييم أداء الدقة للنهج الحالية لتحليل المشاعر، بما في ذلك التعلم الآلي، والتعلم العميق على التغريدات العربية للتعرف على رأي عملاء شركات الاتصالات. تم استخدام الانحدار اللوجستي (LR) وآلة ناقلات الدعم (SVM) والغابة العشوائية (RF) والشبكة العصبية (Neural Network) للكشف عن اتجاه المشاعر لمجموعة البيانات التي تتكون من ١٠٩٨ تغريدة. تشير النتائج إلى أن تقنية التعلم العميق باستخدام طريقة تضمين الكلمة (Word Embeddings)كانت واعدة من حيث الدقة توصلت الى ٨١٪. علاوة على ذلك، في الكشف الجانبي عن الخدمات، أشارت النتائج إلى أن خوارزمية الشبكة العصبية التلافيفية (CNN) لمجموعة البيانات التي تتكون من ١٢٧٧ تغريدة حققت أفضل دقة عند تطبيقها مع أجزاء الكلام (POS) لاستخراج الأسماء التي تتعلق بكل فئة وصلت إلى ٧٥٪. بالإضافة إلى ذلك، في هذه الرسالة، درسنا تحليل المشاعر القائم على الخدمات وأضفنا ميزة جديدة تستخرج اسم الموقع الجغرافي او المكان بناءً

على محتوى التغريدات كمرحلة أخرى. ان استخراج المواقع المكانية للتغريدات سيساعد صانع القرار على اتخاذ أفضل إجراء بناءً على المنطقة أو المدينة أو الحي، والذي بدوره سيساعد الشركة على تحديد المشكلة وموقعها ثم الانتقال إلى حلها. أفاد نموذج التعرف على اسماء الكائنات NER للموقع المقترح الدرجات التالية ٠,٦ و٠,٨٩ من حيث الدقة لنقطة الاهتمام والمدينة NE على التوالي.

# Aspect-based Sentiment Analysis and Location Detection for Arabic Tweets Using Deep Learning
# Case Study: Saudi Telecommunication Companies Tweets

**By:**
**Norah Fahad Ali Alshammari**

**Under supervisor:**
**Dr. Amal Abdullah AlMansour**

## ABSTRACT

In the last few years, Twitter becomes the most popular platform for individuals to share their viewpoints and experiences towards different services and products. Therefore, it attracts a lot of researchers to use it as a body for opinion mining and sentiment analysis. Most of the previous research studies in this area have been using the traditional machine learning-based approach and the lexicon-based approach to classify the emotional states of the tweets in English language. There is limited research work has been done to assort the opinion orientations of tweets in other languages such as Arabic. In addition, recently, deep learning approach has achieved remarkable results over the traditional machine learning algorithms in analyzing a massive amount of data as the case with social networks data. In this thesis, we seek to determine if deep learning approach can be adopted to enhance the sentiment analysis performance for Arabic tweets. Therefore, we evaluated the accuracy performance of existing approaches for sentiment analysis, including machine learning, and deep learning on Arabic tweets to recognize the opinion of the Saudi Telecommunication Companies' customers. Logistic Regression (LR),

Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN) have been utilized to detect the sentiment orientation for a dataset consist of 1098 tweets. Results suggest that deep learning technique with Word Embedding method was promising in terms of accuracy (F1=0.81). Moreover, in the aspect detection, results showed that using CNN algorithm for the dataset contained of 1277 tweets achieved the best accuracy reached to 75% when applied with Part of Speech (POS) features. In addition, in this research work, we added a new feature which is extracting the geographical location based on the tweet content. The proposed location model reported the following scores 0.6 and 0.89 in term of precision for point of interest and city respectively.