Journal of Iranian Architecture & Urbanism
(JIAU)
Homepage: https://www.isau.ir/

ORIGINAL RESEARCH PAPER

A predictive model for analyzing users' perception of the cities based on the Twitter; Case studies: Iran metropolitans

Maryam Mohammadi [1,*]

[1] *Associate Professor, Urban Design Department, Faculty of Urban Planning & Architecture, University of Art, Tehran, Iran/ Visiting Scholar, Statistics and Actuarial Science Department, Faculty of Mathematics, University of Waterloo, Waterloo, Canada.*

ARTICLE INFO

Article History:

Received	2022/06/15
Revised	2022/09/26
Accepted	2023/01/07
Available Online	2023/12/27

Keywords:

Emotion Analysis
Machine Learning
Deep Learning
Twitter
Social Network

Use your device to scan and read the article online

Number of References
46

Number of Figures
3

Number of Tables
4

Extended ABSTRACT

BACKGROUND AND OBJECTIVES: The examination of users' emotions through social media has developed into an impactful domain across diverse scientific fields, appealing not only to business proprietors and politicians but also to general users. In the meantime, this field has infiltrated urban studies and has been used by urban planners and designers due to its methodology; whether in the form of research that aims solely at emotion analysis or as an integrated layer within broader research endeavors. The aim of this article is to explain this field in the analysis of urban emotions as modeling methods in order to identify the position of this field in urban studies by examining the importance of emotion and the methods of its study in the city.

METHODS: This research used the supervised machine learning approach and analyzed the sentiments of tweets related to eight major cities in Iran. The data collection consists of 930 tweets that were collected in a period of 10 years from 2011 to 2022. Initially, over 5000 tweets were collected, and during the tagging process, 80% of them were excluded due to their limited relevance to the city, emphasizing tweets related to urban space. The name of cities and tourist areas were searched to establish a balance between positive and negative data. The tweets are downloaded through Twitter streaming API and the metadata along with the text, including the number of retweets, number of likes and tweet ID, language and location. The data sets have been used for machine training after standard and normalization steps. In this research, the ratio of training data to testing data is 80 to 20. According to the supervised approach, the data were labeled by the researcher with three negative, neutral, and positive labels, and where the researcher had doubts, the opinions of two other experts were used. In general, both machine learning and deep learning have been used. In order to check the validity of the model and to test it, the confusion matrix has been used.

FINDINGS: Firstly, the machine was trained based on 3 algorithms that were used in many research related to text sentiment analysis. Based on the test results presented on the confusion matrix, the accuracy of the trained machine in determining the polarity of the text in three polarities was defined. Among the three used algorithms, support vector machine and random forest have performed better than other algorithms. Given that the model's highest accuracy was approximately 70%, deep learning was employed to train the machine in order to assess the potential for achieving improved results. In the following, machine learning with a convolutional neural network algorithm and a hybrid algorithm were considered. At first, the machine was trained using a convolutional neural network. The results of the accuracy of the model showed that the model is predictable by up to 75%. Next, an attempt was made to improve the predictive accuracy of the model by writing a hybrid algorithm based on the convolutional neural network. The architecture of this network is such that two types of data are considered as input to the neural network, text data and other features in the data set, including location, number of retweets, number of likes, city codes and searched content (as metadata). Therefore, based on this input and output (classification based on the polarity of the text by the researcher), the machine was trained and finally tested. As depicted in the structure of the hybrid algorithm, the significance of the text is assigned a weight of 90%, while the

© 2023, JIAU. All rights reserved.

https://dx.doi.org/10.30475/isau.2023.337531.1905

OPEN ACCESS

* Corresponding Author:
Email: *m.mohammadi@art.ac.ir*
Phone: +98(21)66733401

Extended ABSTRACT

importance of metadata is weighted 10%. It should be noted that different percentages were given to the importance of each of the inputs and the predictability accuracy of the model was checked. As the model test results show, the designed algorithm has improved the predictability of the machine by 4%.

CONCLUSION: In this article, sentiment analysis based on model-oriented methods - machine learning and deep learning - was scrutinized, and therefore, while comparing it with traditional methods and lexical methods, the process of urban sentiment analysis was developed and the different levels of the process were described in detail. As stated, these methods have many advantages and can be useful for analyzing the current situation or predicting different urban projects. Besides, compared to traditional methods, they are less expensive, faster and have sufficient accuracy. According to the appropriate capability of the trained machine, this machine can predict the polarity of the data. This means that by using the text data published in social networks, it is possible to analyze the feelings of users. Certainly, when these data are geodatabases, there is also the capability to geolocate emotions. This approach allows for a swift, accurate, and cost-effective general assessment of city spaces. By identifying areas where users perceive negative emotions, the reasons can be investigated and addressed accordingly. This research has been innovative in two aspects, 1) preparing a collection of data related to the sentiment of Persian language users related to the city and 2) analyzing urban sentiment in the country using machine learning in the field of urban planning and design. Some limitations of this research include limited access to all data published on Twitter using Twitter streaming API; the small amount of available and relevant data; low use of Twitter by users due to filtering; and the unavailability of financial resources to prepare and use a larger set of data.

.....

HIGHLIGHTS:

- In this article, using social networks, a different method for analyzing urban emotions is presented which is based on the modeling method- Machine Learning.
- This method, due to its predictive nature and using Big Data, allows auditing at a lower cost and faster way.
- According to the collected data set, this research prepared the Persian data set related to the city, which indicates the emotional analysis.

ACKNOWLEDGMENTS:
This research did not receive any specific grant from funding agencies in the public, commercial, or not-forprofit sectors.

CONFLICT OF INTEREST:
The authors declared no conflicts of interest.

COPYRIGHTS
Copyright for this article is retained by the author(s), with publication rights granted to the Journal of Iranian Architecture & Urbanism (JIAU). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.
https://creativecommons.org/licenses/by/4.0/

HOW TO CITE THIS ARTICLE
Mohammadi, M., (2023). A predictive model for analyzing users' perception of the cities based on the Twitter; Case studies: Iran metropolitans. *Journal of Iranian Architecture & Urbanism*., 14(2): 301-317.

https://dx.doi.org/10.30475/isau.2023.337531.1905
https://www.isau.ir/article_185079.html

نشریه علمی معماری و شهرسازی ایران. دوره ۱۴. شماره ۲. صفحات ۳۱۷-۳۰۱. پاییز و زمستان ۱۴۰۲

نشریه علمی معماری و شهرسازی ایران
(JIAU)
https://www.isau.ir/

مقاله علمی- پژوهشی

ارائه‌ی مدل پیش‌بینی کننده تحلیل احساسات کاربران از شهر مبتنی بر شبکه‌ی اجتماعی توئیتر؛ نمونه مطالعاتی: کلان‌شهرهای ایران

مریم محمدی [۱و*]

۱. دانشیار، گروه طراحی شهری، دانشکده معماری و شهرسازی، دانشگاه هنر، تهران، ایران/ پژوهشگر، گروه آمار، دانشکده ریاضی، دانشگاه واترلو، واترلو، کانادا.

مشخصات مقاله

تاریخ ارسال ۱۴۰۱/۰۳/۲۵
تاریخ بازنگری ۱۴۰۱/۰۷/۰۴
تاریخ پذیرش ۱۴۰۱/۱۰/۱۷
تاریخ انتشار آنلاین ۱۴۰۲/۱۰/۰۶

واژگان کلیدی

تحلیل احساسات
یادگیری ماشین
یادگیری عمیق
توئیتر
شبکه‌ی اجتماعی

چکیده

تحلیل احساسات کاربران از طریق شبکه‌های مجازی، به‌حوزه‌ای موثر در علوم مختلف تبدیل شده و مخاطبان آن نه تنها صاحبان شرکت‌ها و سیاست‌مدارن، بلکه کاربران هستند. در این میان این حوزه در مطالعات شهری هم نفوذ کرده و به‌دلیل روش‌مندی آن؛ چه در قالب پژوهش‌هایی که صرفاً تحلیل احساس را هدف خود قرار داده‌اند و چه به‌صورت لایه‌ای تلفیقی در پژوهش‌ها مورد استفاده برنامه‌ریزان و طراحان شهری قرار گرفته است. مقاله‌ی پیش‌رو با هدف تبیین این حوزه در تحلیل احساسات شهری در قالب روش‌های مدل‌گرا بر آن است تا با بررسی اهمیت احساس و روش‌های مطرح بررسی آن در شهر، جایگاه این حوزه را در مطالعات شهری نشان دهد و در ادامه به آموزش ماشین برای ارائه‌ی مدل پیش‌بینی‌کننده برای تحلیل احساسات شهر بپردازد. مجموعه‌ی داده‌های این پژوهش مربوط به ۸ کلان‌شهر ایران است که از توئیتر استخراج شده و تحلیل داده‌های متنی مورد توجه قرار گرفته است. به‌منظور آموزش ماشین برای تحلیل احساسات از یادگیری ماشین و یادگیری عمیق بهره برده شده و نتایج آنها با هم مقایسه شده است. الگوریتم‌های مورد استفاده در یادگیری ماشین، ماشین بردار پشتیبان، رگرسیون لجستیک و درخت تصمیم بوده و در یادگیری عمیق، ماشین با استفاده از شبکه‌ی عصبی و شبکه‌ی هیبریدی آموزش و تست شده است. براساس نتایج یادگیری عمیق برای پیش‌بینی احساسات و قطبیت متن در کلان‌شهرهای ایران بهتر عمل کرده و دقتی برابر با ۸۰ داشته است.

نکات شاخص

- در این مقاله، با استفاده از شبکه‌های اجتماعی، منبع جدید و روش مبتنی بر آن برای تحلیل احساسات شهری ارائه شده و به‌جای روش‌های سنتی، از روش مدل‌گرا و فرآیند مبتنی بر آن استفاده شده است.
- این روش به‌دلیل ماهیت پیش‌بینی‌پذیر آن و قابلیت بهره‌مندی از داده‌های کلان، امکان ممیزی با هزینه‌ی کمتر و با سرعت بیشتر را فراهم می‌آورد.
- با توجه به مجموعه‌ی داده‌های جمع‌آوری شده، این پژوهش به تهیه‌ی مجموعه داده فارسی مربوط به شهر که تحلیل احساسات را نشانه می‌رود، دست یافته است.

نحوه ارجاع به مقاله

محمدی، مریم. (۱۴۰۲). ارائه‌ی مدل پیش‌بینی کننده تحلیل احساسات کاربران از شهر مبتنی بر شبکه‌ی اجتماعی توئیتر؛ نمونه مطالعاتی: کلان‌شهرهای ایران، *نشریه علمی معماری و شهرسازی ایران*، ۱۴(۲)، ۳۰۱-۳۱۷.

* نویسنده مسئول
تلفن: ۰۰۹۸۲۱۶۶۷۳۳۴۰۱
پست الکترونیک: m.mohammadi@art.ac.ir

مقدمه

موضوع تحلیل احساسات شهری، یکی از حوزه‌های مهم در فرایند برنامه‌ریزی و طراحی شهری است. از آنجا که فرد احساس خود در ارتباط با محیط را به‌صورت ناخودآگاه، در قالب احساس خوب یا بد و یا ترس، شادی، استرس و... نمایان می‌سازد، این موضوع به یکی از زمینه‌های پژوهشی در شهر تبدیل شده است. چراکه رابطه فرد با محیط را تحت‌تأثیر قرار داده و منجر به جذب یا طرد فرد از فضا می‌شود. در بررسی این موضوع و عوامل مؤثر بر آن روش‌های متعددی تاکنون موردتوجه بوده است. برای مثال پژوهش‌هایی چه در محیط واقعی و چه محیط مجازی از طریق پرسش‌گری و یا از طریق آزمون و بهره مندی از تجهیزات متعدد شناختی، به بررسی تأثیرگذاری عوامل محیطی بر احساس پرداخته‌اند. در این میان تعداد محدودی از پژوهش‌ها، مبتنی بر شاخه تحلیل احساسات شهری و داده‌های شبکه‌های اجتماعی هستند. این حوزه که به‌صورت کلی مبتنی بر دو روش مبتنی بر قانون و روشهای مدلگرا است، از علم زبانشناسی، آمار و کامپیوتر بهره می‌گیرد و به تحلیل احساسات از حیث قطبیت یا نوع احساس می‌پردازد. پژوهش‌های بسیاری در حوزه‌های مختلف سیاسی، اجتماعی، مدیریتی، اقتصادی و ... از طریق تحلیل احساسات انجام شده که منجر به تحلیل وضع موجود و یا ارائه مدل‌های پیش‌بینی‌کننده شده است.

این پژوهش بر آن است تا ضمن معرفی این نوع از تحلیل و روش‌های آن، با ارائه‌ی فرایند انجام آن مبتنی بر روش‌های مدلگرا (یادگیری ماشین) به ارائه‌ی نحوه‌ی کاربست آن در محیط شهری بپردازد. از آنجا که روش‌های مدلگرا که مبتنی بر هوش مصنوعی هستند قابلیت مناسبی برای ارائه‌ی مدل‌های پیش‌بینی‌کننده را دارند، این پژوهش مترصد است تا فرایند انجام تحلیل احساسات را با توجه به نوع مسئله که از نوع طبقه‌بندی است، تبیین کند. این پژوهش صرفاً به تبیین کاربرد و تدوین مدل مبتنی بر این روش بسنده نکرده و در ادامه در این پژوهش تلاش شده، تحلیل احساسات محیط شهری از طریق رویکرد مدل‌گرا در شهرهای ایران پرداخته شود. بدین ترتیب ضمن تهیه مجموعه داده‌های شهری مربوط به احساسات شهری، با آموزش ماشین، یک مدل پیش‌بینی‌کننده احساسات در شهر نیز ارائه شده است.

مبانی نظری

در این بخش در ابتدا تعریفی از احساس و اهمیت آن در مطالعات شهری ارائه و سپس روش تحلیل احساسات در شبکه‌های اجتماعی شرح داده شده و در نهایت نحوه‌ی کاربرد آن با مرور تجارب این حوزه ارائه می‌شود.

احساس از نظرگاه بیولوژیکی

احساس انتقال اثر محرک (یا پیام فرستنده) از گیرنده‌ی حسی به سیستم اعصاب مرکزی است، امری که به‌صورت عینی قابل پیگیری است (Pakzad, 2006, 3). بورینگ[1] معتقد است که احساس اگرچه زاییده‌ی دریافت‌های جسمی ما از طریق حواس پنج‌گانه است؛ اما چیزی است غیرمادی و صرفاً روحی (Grutter, 1987, 7; Pakzad, 2006,149). احساس، ارزش شناختی نداشته و به‌گونه‌ای منفعل حاصل می‌شود و صددرصد فیزیولوژیک است (Pakzad, 2006, 3).

محیط بر رفتار، احساسات و حالات افراد تأثیر می‌گذارد. مکان‌ها می‌توانند به‌عنوان کسل‌کننده[2]، هیجانی[3]، ترسناک[4]، تسلی‌بخش[5]، افسرده کننده[6] و یا متعالی[7] تجربه شوند. کیفیت‌های عاطفی مکان‌ها می‌توانند انسان‌ها را جذب نموده یا آنها را طرد و مضطرب نماید (Houtkamp, 2012, 16). راسل و اسنودگرس (۱۹۸۷) نیز بیان می‌کنند که اهمیت کیفیت‌های عاطفی مکان‌ها در توصیفات انسان‌ها از محیط‌هایی که در آن زندگی، کار و تفریح می‌کنند واضح است؛ فضاها همواره برچسب‌های بسیاری مانند ملال آوری[8]، افسردگی[9]، گشودگی[10]، الهام بخش بودن[11]، آرامش[12]، برانگیزاننده[13]، صلح‌آور [14] و ... را با خود حمل می‌کنند (Houtkamp, 2012, 22). برخی از مطالعات در زمینه روانشناسی محیطی نشان می‌دهند که انسان‌ها به طور عاطفی اطراف خود را ادراک و ارزیابی می‌نمایند. برخی مکان‌ها به‌عنوان محلی نا امن تجربه می‌شوند، درحالی‌که برخی دیگر جالب و جذاب ارزیابی می‌شوند. تجربیات زندگی روزانه نشان می‌دهد که بسیاری از رفتارهای روزانه و تصمیم‌گیری‌های فرد، غالباً از این نوع پاسخ‌های عاطفی محیط تأثیر می‌پذیرد (Huang et al., 2014, 93). بنابراین احساسات کاربران در شهر بر رفتار و عمل آنها در شهر تأثیر دارد، از همین رو تحلیل احساسات در شهر به یکی از حوزه‌های مهم پژوهشی تبدیل شده است.

در پژوهش‌های مختلف، برای سنجش احساسات از روش‌های مختلفی استفاده شده است. براساس پژوهش‌ها اندازه گیری احساسات از سه طریق انجام می‌شود: ۱) رسانه‌ی اجتماعی؛ ۲) اپلیکیشن‌ها؛ و ۳) دستگاه‌های پوشیدنی (Fathullah & S.Willis, 2018, 5). ازآنجاکه هدف این مقاله نحوه‌ی کاربست رسانه‌های اجتماعی در تحلیل احساسات شهری است، به بررسی آن پرداخته می‌شود.

تحلیل احساسات از طریق شبکه‌های اجتماعی

تحلیل احساسات یکی از شاخه‌های متن‌کاوی است، هر چند این شاخه در داده‌های غیرمتنی نیز رواج یافته است. این تحلیل، زمینه‌ی مطالعاتی مربوط به تحلیل نظرات، نگرش‌ها و احساسات افراد نسبت به موجودیت‌هایی مانند محصولات، خدمات،

سازمان‌ها، مکان‌ها و رویدادها است (Liu, 2015). تحلیل احساس، به معنی فهم احساسات مثبت و منفی کاربر از طریق پست‌های انتشار یافته در شبکه‌های اجتماعی است (Sim et al., 2020). تحلیل احساسات با تکیه بر قطبیت یک کلمه به بررسی جمله یا سند می‌پردازد، یعنی آیا متن، احساس مثبت، منفی یا خنثی را منتقل می‌کند (Liu & Zhang, 2012 به نقل از Bonasoli de Oliveira et al., 2020). این نوع از تحلیل، یک روش تحلیل شبه کیفی است؛ شکلی از تحلیل محتوا که می‌تواند برای کلان داده‌ها مثل مجموعه‌ی داده‌های رسانه‌های اجتماعی استفاده شود(Hollander & Renski, 2015).

جایگاه شبکه‌های اجتماعی در تحلیل احساسات

در گذشته نظرکاوی از طریق برگه‌های نظرسنجی و یا آزمایش بر روی یک گروه خاص انجام می‌شد، اما امروزه با گسترش روزافزون رسانه‌های اجتماعی، محتوای مطالب منتشر شده در رسانه‌ها برای تصمیم‌گیری مورد استفاده قرار می‌گیرد (Shah Talebi et al., 2015). در گذشته براساس تحلیل محتوای سنتی، محقق سند را می‌خواند و کلمات و عبارات خاصی را کدگذاری می‌کرد، اما در روش نوین، این فرایند خودکار بوده و با استفاده از فرهنگ لغت احساسات و برنامه‌ی رایانه‌ای انجام می‌شود (Renski, 2015).

رسانه‌های اجتماعی به دلیل ظهور ویژگی‌های «وب ۲,۰» و توسعه‌ی دسترسی به اینترنت و افزایش استفاده از تلفن‌های هوشمند رشد چشم‌گیری داشته‌اند (Lin & Geertman, 2019; Kitchin, 2014). به نقل از Sim et al., 2020). وب ۲,۰ مجموعه‌ای از فناوری‌ها است که به کاربران امکان می‌دهد، محتوا تولید کنند، با دیگر کاربران ارتباط برقرار کرده و اطلاعات را به اشتراک بگذارند (Kaplan & Haenlein, 2010 به نقل از Sim et al., 2020). تفاوت این وب از وب سایت‌های کلاسیک و وبلاگ این است که محتوا توسط افراد خاصی نشر نمی‌شود، بلکه تمامی کاربران به طور مداوم در حال تغییر و توسعه آن هستند. استرن (۲۰۱۰)، رسانه‌های اجتماعی را به هفت دسته با توجه به گستره و محتوا تقسیم می‌کند (Sterne, 2010 به نقل از Sdoukopoulos et al., 2018): ۱) فورم‌ها: جهت تبادل افکار، نظرات و تجربیات؛ ۲) سایت‌های نقد و نظر: اشتراک نظرات و پاسخ‌دهی به بازخوردها به‌عنوان مثال یلپ و تریپ ادوایزر[۱۵]؛ ۳) شبکه‌های اجتماعی: بستر ارتباط افراد و اشتراک علایق مانند فیس‌بوک و لینکدین؛ ۴) میکرو وبلاگینگ: همانند وبلاگنویسی که از پست‌های کوتاه پشتیبانی می‌کند (مانند توئیتر)؛ ۵) نشانک‌گذاری: ذخیره، سازماندهی و به اشتراک گذاری اینترنتی با سایر کاربران؛ ۶) اشتراک‌گذاری رسانه: بارگذاری و اشتراک‌گذاری تصاویر و فیلم مانند یوتوب و فلیکر[۱۶]. کاپلان و هنلین نیز رسانه‌های اجتماعی را مشتمل بر این دسته‌ها می‌دانند (Kaplan & Haenlein, 2010 به نقل از Sim et al., 2020): نوع اول رسانه‌های مشارکتی مانند ویکی‌پدیا هستند؛ نوع دوم برای اشتراک گذاری محتوا کاربرد دارند مانند یوتوب و فلیکر؛ نوع سوم، پلتفرم‌های شبکه‌های اجتماعی مانند توئیتر و فیس‌بوک هستند؛ و نوع چهارم شامل پلتفرم‌های وبلاگنویسی مانند وردپرس و تیومبلر[۱۷] هستند.

نقاط قوت و ضعف بهره‌گیری از داده‌های شبکه‌های اجتماعی برای تحلیل احساسات

تحلیل داده‌های رسانه‌های اجتماعی و میکرو وبلاگ به‌عنوان وسیله‌ای برای جمع‌آوری اطلاعات در مورد مسائل اجتماعی دارای نقاط قوت و ضعف هستند. این روش، یک روش نسبتاً سریع و کم هزینه برای جمع‌آوری نظرات در زمان واقعی از طیف وسیعی از مردم و در رابطه با موضوعات متعدد است. این روش بسیار ساده‌تر، ارزانتر و سریع‌تر از انجام نظرسنجی یا مصاحبه است. با این حال، محدودیت‌هایی نیز وجود دارد، برای مثال استفاده از رسانه‌های اجتماعی برای ابراز عقاید و احساسات در بین گروه‌های سنی خاص و در میان افرادی که دسترسی بیشتری به تلفن‌های هوشمند و رایانه نسبت به سایر گروه‌ها دارند، بسیار رایج‌تر است (Hollander & Renski, 2015).

گزارش مرکز تحقیقات پیو در سال ۲۰۱۲ نشان داد که تنها ۱۵ درصد از بزرگسالان در ایالات متحده از توئیتر استفاده می‌کنند و این افراد بین ۱۸ تا ۲۹ سال سن دارند و در مناطق شهری زندگی می‌کنند (Smith & Brenner, 2012 به نقل از Hollander & Renski, 2015). علاوه بر این مشخص شد که کاربران توئیتر به‌طور قابل‌توجهی بیشتر مرد بوده و در مناطق پر جمعیت زندگی می‌کنند. در عین حال عوامل اجتماعی، اقتصادی، زبانی و فرهنگی وجود دارند که ممکن است بر استفاده از رسانه‌های اجتماعی نیز تأثیرگذارند. بنابراین نمی‌توان گفت که همه رسانه‌های اجتماعی یا میکروبلاگ‌ها و داده‌های مستخرج از آن یک نمونه‌ی تصادفی مناسب از جمعیت برای مطالعه است (Hollander et al., 2014 به نقل از Hollander & Renski, 2015). برخی دیگر بر این باور هستند که این روش تحلیل، به‌دلیل مختصر بودن پست‌ها و تمایل به استفاده از کلمات اختصاری می‌تواند چالش‌برانگیز باشد (Zhai & Massung, 2016به نقل از Sim et al., 2020).

با توجه به کاربرد زیاد تحلیل احساسات شهری براساس داده‌های توئیتر، به‌صورت مجزا، اهمیت، مزایا و نمونه پژوهش‌های انجام شده با استفاده از توئیتر بررسی می‌شود. یکی از مزایای توئیت‌ها در تحلیل احساسات، تکرار، سرعت و دقت آنها در انتقال احساسات است (Roberts et al., 2018, 22). در ادامه برخی از مزایای کاربرد توئیتر ارائه می‌شود:

- کاربران توئیتر پیام‌های کوتاه تولید می‌کنند، محدود به ۲۸۰ کاراکتر؛ البته تا اواخر سال ۲۰۱۷، محدودیت کاراکتر ۱۴۰ بوده است؛
- راحتی و دسترسی به توئیتر و جمع‌آوری داده‌ها؛
- تعدد پژوهش‌هایی که از این منبع به‌عنوان منبع اصلی استفاده نموده‌اند (Sim et al., 2020).

از دیگر مزایای ذکر شده برای استفاده از توئیتر در تحلیل داده‌ها می‌توان به موارد زیر اشاره نمود:

- توئیتر چهاردهمین وبسایت پر بازدید در سراسر جهان در سال ۲۰۱۶ بوده است؛
- کاربران از طریق توئیتر افکار و نظرات خود را در مورد مسائل مختلف به اشتراک می‌گذارند؛
- توئیتر با سایت‌های میکروبلاگینگ مطابقت دارد و آن را به بستری مناسب برای تبادل نظرات و ایده‌های کوتاه تبدیل می‌کند؛

آسان بودن جمع‌آوری و طبقه‌بندی داده‌های مربوط به یک موضوع یا موضوع خاص نسبت به سایر رسانه‌های اجتماعی، ازآنجایی‌که از هشتگ استفاده می‌شود؛

- ای.پی.آی. توئیتر در مقایسه با سایر رسانه‌های اجتماعی مانند فیس‌بوک بازتر و قابل دسترس‌تر است و به برنامه نویسان اجازه می‌دهد تا به حجم زیادی از داده‌ها دسترسی داشته باشند؛
- به دلیل ابزارهای ساده و خودکار استخراج داده در توئیتر مانند NodeXL، توئیتر بسیار کاربردی و کاربر پسند است و در ذخیره زمان نیز مؤثرتر است (Al-exa Internet Inc., 2017 به نقل از Sdoukopoulos et al., 2018).

علی‌رغم مزایا و کاربردهای توئیتر در تحلیل احساسات شهری، مشکلاتی نیز وجود دارد. برای مثال کار با این نوع از محتوای متنی غیررسمی، چالش‌های جدیدی را برای پردازش زبان ایجاد می‌کند؛ زیرا زبان مورد استفاده در توئیتر اغلب با کاربرد علائم نگارشی و املایی عامیانه و اختصاری همراه است (Rosenthal et al., 2014 به نقل از Rosenthal et al., 2014). همچنین کاربرد ایموجی‌ها از دیگر چالش‌های تحلیل‌گران است.

انواع داده‌های شبکه اجتماعی برای تحلیل احساسات

در متن‌کاوی، تحلیل احساسات به‌عنوان یک شاخه بیان شده، اما پژوهش‌هایی مرتبط با تحلیل احساسات داده‌های غیرمتنی نیز رواج یافته است.

داده‌های متنی: یکی از متداولترین نوع داده‌ها در شبکه‌های اجتماعی، داده‌های متنی است و پژوهش‌ها و توسعه زیادی در زبان‌های مختلف برای تحلیل احساسات انجام شده است.

داده‌های بصری: فارغ از بررسی متون برای تحلیل احساسات، امروزه تحلیل احساسات از طریق تصاویر نیز مورد توجه قرار گرفته (Campos et al., 2017; Chen et al., 2014; Dhall et al., 2015)؛ چرا که اشتراک‌گذاری تصاویر برای بیان حالات و احساسات بسیار رایج شده است (Parret, 2016; You et al., 2015). به نظر می‌رسد که توسعه‌ی تکنیک‌های جدید برای این منظور می‌تواند رویکردهای مبتنی بر متن را تکمیل کند (Zhan et al., 2019; Zhou & Xiangji Huang, 2017; Riberio et al., 2016; Benevenuto et al., 2015 به نقل از Bonasoli de Oliveira et al., 2020).

داده‌های چند وجهی: تحلیل داده‌های چندوجهی، شامل داده‌های متنی و تصویری (setaiawan et al., 2021; ji et al., 2015) و یا داده‌های متنی، تصویری و صوتی (Poria et al., 2017) نیز به عنوان رویکردی جدید در برخی از پژوهش‌ها مورد توجه قرار گرفته است.

روش‌های تحلیل احساسات

رویه‌های مختلف تا کنون برای تحلیل احساسات ارائه شده؛ برخی از آنها بر مبنای زبان شناسی رایانشی طراحی شده و بیشتر بر پایه یادگیری ماشین پایه‌ریزی شده‌اند و بر مبنای دیدگاه پنگ[18] هستند که تحلیل احساس و طبقه بندی متون را بر اساس سه روش یادگیری ماشین نظارت شده شامل بیز ساده، آنتروپی بیشینه و ماشین بردار[19] پشتیبانی انجام داده‌اند (Pourmostafa Roshan Sharemi et al., 2017 به نقل از Pang et al., 2002). در دسته‌بندی دیگری تکنیک‌های متفاوتی در این ارتباط بیان شده که عبارتند از: ۱) یادگیری ماشین و یادگیری عمیق؛ و ۲) تکنیک مبتنی بر واژه که خود شامل تکنیک‌های مبتنی بر قانون؛ تکنیک مبتنی بر فرهنگ واژه؛ تکنیک مبتنی بر رابطه و تکرار و تکنیک‌های ترکیبی می‌شود (Sabohi, 2018 به نقل از Devika et al., 2016). در ادامه شرحی بر این روش‌ها آورده می‌شود.

روش‌های مبتنی بر واژه یا روش لغوی[20]

انواع روش‌هایی که ذیل روش لغوی طرح شده‌اند، به شرح زیر هستند:

روش مبتنی بر قانون: این تکنیک با تعریف قوانین مختلف برای تحلیل عقاید با استفاده از نشانه‌گذاری هر جمله در هر سند و سپس تست هر علامت یا کلمه پیش می‌رود. هر پست در ابتدا امتیاز صفر دارد، اگر کلمه‌ای با احساسات مثبت وجود داشته باشد، امتیاز ۱+ به آن اعمال و اگر منفی باشد، امتیاز ۱- در نظر گرفته می‌شود (Devika et al., 2016 به نقل از Yousofi Moteghaed & Sabohi, 2018).

روش مبتنی بر واژه‌نامه‌ها: این تکنیک مبتنی بر واژه‌نامه است که قطبیت کلی یک جمله یا سند برابر با مجموع قطبیت‌های عبارات و کلمات موجود در آن است. این روش برای آنالیز احساسی هر واژه است. تکنیک مبتنی بر قانون نیز بر پایه فرهنگ

لغت کار می‌کند (Devika et al., 2016 به نقل از Yousofi Moteghaed & Sabohi, 2018).

روش‌های مبتنی بر تکرار و رابطه: بیشتر روش‌هایی که در تحلیل احساسات در سطح ویژگی استفاده شده‌اند، مبتنی بر روش‌های تکرار هستند (Zhu et al., 2011; Hu & Liu, 2004; Wei et al., 2010; Popesc & Etzioni, 2005; Mei et al., 2007 به نقل از Shah Talebi et al., 2015). این روش‌ها معمولاً محدودیت‌هایی را بر روی عبارات اسمی با تعداد تکرار بالا برای شناسایی ویژگی‌ها و احساسات در نظر می‌گیرند. درحالی‌که روش‌های وابسته به رابطه از روابط بین کلمات در متن استفاده می‌کنند تا ویژگی و احساسات را تشخیص دهند (Qiu et al., 2011; Guo et al., 2009; Liu et al., 2005; Zhuang et al., 2006 به نقل از Shah Talebi et al., 2015). پژوهش‌هایی نیز در رابطه با ترکیب این دو روش وجود دارد (Moghaddam & Ester, 2010 ; Zhang et al., 2010 به نقل از Shah Talebi et al., 2015).

روش‌های مدل‌گرا - یادگیری ماشین و یادگیری عمیق در تحلیل احساسات یادگیری ماشین: یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به سیستم‌های رایانه‌ای امکان یادگیری مستقیم از مثال‌ها، داده‌ها و تجارب را می‌دهد. سیستم‌های یادگیری ماشین می‌توانند فرایندهای پیچیده‌ای را با یادگیری از داده‌ها، به‌جای پیروی قوانین از پیش برنامه‌ریزی شده، انجام دهند (Royal Society (Great Britain), 2017, 5). یادگیری ماشین توسط دانشگاه استنفورد[۲۱] به‌عنوان «علم وادارکردن رایانه‌ها بدون برنامه‌ریزی صریح[۲۲]» تعریف شده است (Keith, 2019, 1). جودیت هورویتز[۲۳] و دانیل کرش[۲۴] (۲۰۱۸) بیان کرده‌اند که یادگیری ماشین سیستم را قادر می‌سازد تا از طریق برنامه‌نویسی صریح از داده‌ها یاد بگیرد. یادگیری ماشین از الگوریتم‌های مختلفی استفاده می‌کند که به طور مداوم از داده‌ها برای بهبود، توصیف و پیش‌بینی نتایج استفاده می‌کند. از آنجا که الگوریتم‌ها (ماشین بردار پشتیبان (Ayodele, 2010)؛ درخت تصمیم‌گیری (Safavian & (Landgrebe, 1991؛ جنگل تصادفی (Breiman, 2001)؛ رگرسیون لجستیک (Malhotra &Semwal, 2019)؛ آدابوست (Hastie et al., 2009) و ایکس جی بوست (Chen & Gues-trin, 2016) و... داده‌های آموزشی را جذب می‌کنند؛ می‌توان مدل‌های دقیق‌تری را بر اساس آنها تولید کرد. تکنیک‌های یادگیری ماشین، در ابتدا الگوریتم را با برخی از ورودی‌های خاص با خروجی شناخته شده آموزش می‌دهند تا سپس بتوانند با داده‌های ناشناخته جدید کار کنند (Huwitz & Kirsch, 2018, 4). الگوریتم‌های احتمالاتی و یادگیری ماشین که از قوانین ریاضی استفاده می‌کنند در حوزه پردازش زبان طبیعی عملکرد مناسبی دارند. این الگوریتم‌ها درصورتی‌که امکان دسترسی به میزان بالای داده وجود نداشته باشد، مورد استفاده قرار می‌گیرند.

در رویکرد یادگیری ماشین، تحلیل احساس با استفاده از قواعد زبانی یا نحوی و یا هر دوی آنها و به‌عنوان یک مسئله‌ی طبقه‌بندی متن عادی تلقی می‌شود و الگوریتم‌های یادگیری ماشین بر روی آن اعمال می‌شود. از میان الگوریتم‌ها، مواردی چون بیز ساده (Pourmostafa Roshan Sharemi et al., 2017 به نقل از Thorsten, 1997)، گرادیان کاهشی تصادفی (Pourmostafa Roshan Sharemi et al., 2017 به نقل از Prasetijo et al., 2017) و ماشین بردار پشتیبانی (Pourmostafa Roshan Sharemi et al., 2017 به نقل از Li & Li, 2013) در زمینه طبقه‌بندی متن به‌خوبی عمل کرده‌اند.

یادگیری عمیق: یادگیری عمیق همان یادگیری ماشین در سطحی عمیق‌تر است. برخلاف الگوریتم‌های یادگیری ماشین استاندارد که مسائل را به بخش‌های کوچک‌تر تقسیم و سپس آنها را حل می‌کنند، یادگیری عمیق مسائل را به شکل کامل حل می‌کند. هر چه حجم داده بیشتر و زمان بیشتری در اختیار الگوریتم‌های یادگیری قرار گیرد، نتیجه بهتر است (Yousofi Moteghaed & Sab-ohi, 2018 به نقل از Devika et al., 2016). یادگیری عمیق در بسیاری از شاخه‌های حوزه پردازش زبان طبیعی همچون تحلیل احساسات، ترجمه ماشین، بینایی ماشین و... کاربرد دارد (Pourmostafa Ros-han Sharemi, et al., 2017 به نقل از LeCun et al., 2015). یادگیری عمیق که در حل بسیاری از مسائل پردازش زبان طبیعی مورد استفاده قرار می‌گیرد، یک مزیت مهم و اصلی دارد. اصلی‌ترین مزیت این رویکرد آن است که نیاز به تنظیم دستی و استخراج مشخصه‌ها براساس دانش تخصصی و دسترسی به منابع زبانی نیست (Pourmostafa Roshan Sharemi, et al., 2017 به نقل از Rojas-Barahona & Ma-ria, 2016). سه ویژگی که سبب برتری روش‌های یادگیری عمیق نسبت به روش‌های سنتی شده عبارتند از: ۱) توانایی این روش‌ها در یادگیری مستقیم از داده‌های خام؛ ۲) داشتن ساختاری سلسله مراتبی و عمیق؛ و ۳) عمومی و بهینه بودن روش نسبت به روش‌های سنتی (Mousavi et al., 2019).

کاربست روش‌های تحلیل احساسات مبتنی بر شبکه‌های اجتماعی در شهر با مرور تجارب

یکی از روش‌های کاربرد تحلیل احساسات در شهر که در پژوهش‌های شهری مورد استفاده قرار گرفته، تحلیل احساسات کاربران در مقیاس فضاهای شهری و شهر است. در ادامه نمونه‌ای از آنها ارائه می‌شود و از خلال بررسی روش و فرایند آنها، آموزه‌های روش شناختی مربوط به استفاده از این روش تحلیل نیز تبیین می‌شود.

در پژوهشی (Roberts et al., 2018) که به‌منظور سنجش احساسات شهری در فضاهای شهری در

بیرمنگام در سال ۲۰۱۴ انجام شد، توئیت‌های مربوط به ۶۰ فضای سبز (۴۶ پارک، ۱۴ فضای خطی شامل مسیرهای پیاده روی در کنار ۴ رودخانه، ۷ کانال و ۳ مسیر دوچرخه‌سواری) به‌عنوان نمونه‌های موردی پژوهش انتخاب شدند. در مرحله جمع‌آوری و آماده سازی داده‌ها از داده‌های توئیتر بهره برده شد و از طریق رست ای.پی.آی.[۲۵] خروجی ارائه و زبان توئیت‌ها صرفاً انگلیسی انتخاب شد. در حین آماده سازی برای جمع‌آوری داده‌های توئیتر، از مقیاس‌های مختلف زمانی برای جمع‌آوری توئیت‌ها با هدف حصول اطمینان از موثرترین فراوانی جمع‌آوری توئیت‌ها استفاده شد. براساس بررسی‌ها، نتیجه نشان داده که استفاده از فراوانی ۳، ۵ و ۷ روز برای دانلود توئیت‌ها منجر به تکراری شدن بیش از حد داده‌ها می‌شود و لذا با انتخاب فراوانی ۱۰ روزه از آنجا که هیچ توئیتی در مقایسه با جستجوی ۷ روز کم نشده، این بازه زمانی مناسب است. از ژوئن ۲۰۱۵ تا ماه مه ۲۰۱۶ توئیت‌ها جمع‌آوری و هر گونه توئیت تکراری در حین پردازش حذف و در مجموع ۱۰۲۶۸ توئیت برای استفاده در مطالعه، ملاک عمل قرار گرفت. در مرحله برچسب‌زنی و حاشیه‌نویسی، از هر سه نوع روش دستی، خودکار و نیمه‌خودکار استفاده و نتایج مدل با هم مقایسه شده‌اند. در روش خودکار[۲۶] از منبع ای.ان.ایی.دبلیو.[۲۷] به‌عنوان ابزاری برای حاشیه‌نویسی احساسات استفاده شد. این منبع از ورینر و همکاران ((۲۰۱۳ گرفته شده که به بیش از ۱۳ هزار کلمه انگلیسی، نمره‌های ظرفیت[۲۸] اختصاص داده است. در روش نیمه خودکار حاشیه‌نویسی (Resch et al., 2016) مبتنی بر نمودار است، نمونه‌ای از هزار توئیت به‌صورت دستی حاشیه‌نویسی شد و براس آموزش و ارزیابی الگوریتم مورد استفاده قرار گرفت.

نمونه دیگری از تحلیل‌های انجام شده در فضاهای شهری در مقیاس خردتر بوده و صرفاً به بررسی دو یا چند فضا پرداخته‌اند. برای مثال در پژوهشی (Sim et al., 2020) که به بررسی تحلیل احساسات در پارک‌های شهری در شهر شیکاگو امریکا پرداخته، تحلیل احساسات در دو پارک پرطرفدار در شیکاگو با نام‌های های‌لاین [۲۹] در نیویورک و ۶۰۶ براساس توئیت‌های مرتبط از سال ۲۰۱۵ تا ۲۰۱۹ انجام شده است. این پژوهش داده‌ها از نوع متنی بوده و از شبکه اجتماعی توئیتر استفاده شده است. پس از پیش‌پردازش متن، با توجه به تحلیل‌های سه‌گانه مشتمل بر تحلیل موضوع، تحلیل احساسات و تحلیل متن، الگوریتم‌ها انتخاب شده‌اند. تحلیل موضوعات برای کشف موضوعات اصلی در مورد هر پارک به صورت ماهیانه انجام شده است. برای تحلیل احساسات، از مدل وی.ای.دی.ایی.آر.[۳۰] استفاده شده که برای داده‌های رسانه‌های اجتماعی توسعه داده شده است (Hutto & Gibert, 2019). این مدل هم واجد دقت است و هم برای متون کوتاه، در مقایسه با سایر مدل‌های تحلیل احساسات مناسب‌تر عمل می‌کند. تحلیل متن، در سطح ابتدایی‌تر، فراوانی کلمات کلیدی منتخب و رابطه‌ی بین آنها را بررسی می‌کند. کلمات کلیدی مربوط به فعالیت‌های بدنی، فعالیت‌های اجتماعی، طبیعت و مناظر را بر اساس پیمایش انتخاب شده‌اند. در ادامه احساسات هر فعالیت نیز شناسایی و مقایسه شده‌اند (Sim et al., 2020). با این توضیح این روش صرفاً به تحلیل احساسات نپرداخته، بلکه از طریق متن‌کاوی، در ابتدا به تحلیل موضوعات پرداخته و از طریق شناسایی فعالیت‌ها، احساسات کاربران را نسبت به نوع فعالیتی که در پارک وجود دارد سنجیده است.

در پژوهش دیگری (Saeidi et al., 2016) از مجموعه داده‌های سنتی‌هود که پلتفرمی برای پرسش و پاسخ است در محلات شهری لندن استفاده شده است. در این پژوهش، موجودیت‌های موجود در مجموعه داده‌ها، مکان‌ها یا محله‌ها هستند. پاسخ‌ها با استفاده از نام هر محله از شهر لندن در نظر گرفته شد. نام‌های مکان از روزنامه جئونیمز[۳۱] گرفته شده است. فقط جملاتی برای تحلیل در نظر گرفته شد که دارای نام مکان هستند و جملات دیگر کنار گذارده شد. تعداد مکان‌های ذکر شده در یک جمله در مجموعه داده از یک تا بیش از ۵۰ مکان، متغیر است. برای ساده‌کردن کار، فقط جملاتی حاشیه نویسی شدند که حاوی یک یا دو مکان هستند. مانند پژوهش‌های موجود در رابطه با تحلیل احساسات مبتنی بر وجه[۳۲] (BrychcIn et al., 2014 به نقل از Saeidi et al., 2016)، لیستی از پیش تعریف شده از وجه‌ها به حاشیه‌نویسان ارائه شده که می‌توانند از بین آنها انتخاب کنند. این وجه‌ها عبارتند از: زندگی، امنیت، قیمت، آرام، ناهارخوری، زندگی شبانه، حمل‌ونقل، گردشگری، خرید، فرهنگ سبز و چند فرهنگی. افزودن وجه نیز در نظر گرفته شد. برای هر وجه انتخاب شده، حاشیه‌نویس‌ها باید قطبیت متن یا احساس متن را شناسایی (مثبت و منفی) کنند. ۱۰ درصد از کل مجموعه داده به طور تصادفی انتخاب و توسط هر سه حاشیه‌نویس، حاشیه‌نویسی شد. حاشیه‌نویس با بالاترین توافق بین حاشیه‌نویس برای حاشیه‌نویسی تمام مجموعه داده انتخاب شد. مجموعه داده این پژوهش همانطور که بیان شد، سنتی هود است. سنتی هود شامل ۵۲۱۵ جمله با ۳۸۶۲ جمله، حاوی یک مکان واحد و ۱۳۵۳ جمله، حاوی چندین (دو) مکان است. در این پژوهش برچسب «مثبت» برای وجه‌هایی مانند ناهارخوری و خرید غالب است. در این میان، وجه عمومی با بیش از ۲۰۰۰ جمله، رایج‌ترین وجه است. درحالی‌که وجه توریستی در کمتر از ۱۰۰ جمله به‌کار رفته است. ازآنجاکه هر جمله می تواند یک یا چند نظر داشته باشد، تعداد کل نظرات (۵۹۲۰) در مجموعه داده بیشتر از تعداد جملات است. در این پژوهش چهار وجه متداول از مجموعه داده انتخاب شدند که عبارتند از قیمت، ایمنی، موقعیت حمل

و نقل و عمومی. در این پژوهش از الگوریتم حافظه کوتاه مدت طولانی[۳۳] استفاده شده است.

تحلیل احساسات در مقیاس شهر نیز مورد تأکید بوده است. برای مثال در پژوهشی (Hollander & Renski, 2015)، دو مجموعه ۵۰ تایی از شهرهای روبه‌زوال و شهرهای درحال رشد (پایدار) مورد بررسی قرار گرفته‌اند. مجموع ۳۰۰ هزار توئیت در ۲ ماه جمع‌آوری شده و احساسات به صورت مثبت و منفی در این دو دسته شهر تحلیل شده‌اند. پس از اجرای آزمون‌های تفاوت میانگین، مشخص شده که احساسات در شهرهای رو به افول از نظر آماری تفاوت معناداری با شهرهای پایدار و در حال رشد ندارد. در این پروژه برای تحلیل احساسات شهری از یک ابزار داده کاوی و تحلیل متن که در سال ۲۰۱۳ با حمایت لینکلن توسعه داده شده و در پروژه تحقیقاتی استفاده شده بهره برد شده که از عملیات زیر پشتیبانی می‌کند: ۱) تحلیل انواع و تعداد زیاد فایل‌های متنی و ۲) دانلودکردن توئیت‌ها از توئیت‌های فیلتر شده بر اساس مکان.

گام اول برای دستیابی به مجموعه داده‌ها، دانلودکردن توئیت‌ها است. گام بعد تحلیل احساسات هر توئیت و برچسب‌زنی داده‌ها است. در اینجا برچسب‌زنی از روش خودکار و مبتنی بر فرهنگ لغت بهره برده است. برای این کار کلمات کلیدی تعریف شده در فرهنگ لغت، با توجه به احساسات آنها با یک عدد صحیح، در نظر گرفته می‌شوند. این برنامه دارای یک فرهنگ لغت پیش‌فرض بر اساس ای.اف.آی.ان.ان.[۳۴] است. این فرهنگ لغت توسط فین اروپ نیلسون[۳۵] توسعه داده شده و کلمات را در مقیاس ترتیبی از ۵- تا ۵+ طبقه‌بندی می‌کند. به‌عنوان مثال، نمره «سوء استفاده کننده» ۳- است، در حالی‌که «نمره رضایت» ۲+ است. آخرین نسخه‌ی این فرهنگ لغت[۳۶]، دارای ۲۴۷۷ کلمه است و قادر بوده کلمات صحیح را شناسایی نماید. برای مثال انواعی از کلمات مانند «woooow» را به‌عنوان «wow» را تشخیص می‌دهد. امتیاز هر کلمه‌ی احساسی برای هر توئیت جمع می‌شود و نمره خالص، معیاری برای احساسات موجود در مجموعه‌ی داده را نشان می‌دهد. بنابراین در این پژوهش از روش‌های لغوی در تحلیل احساسات بهره برده شده و براساس آنها میانگین امتیاز هر توئیت مشخص شده و قطبیت هر توئیت نیز تعیین شده است.

همانطور که بیان شد تحلیل احساسات در شبکه‌های اجتماعی صرفاً متکی بر متن نبوده و در رابطه با شهر از داده‌های تصویری نیز استفاده شده است؛ برای مثال پژوهشی که در سال ۲۰۲۰ به تحلیل احساسات تصویری در فضا های باز شهری در شهر شیکاگو در آمریکا پرداخت. در این پژوهش، نظرات کاربران به استناد به تصاویر ارسال شده از فضاهای باز شهری در حین فعالیت مورد بررسی قرار گرفت (Bonasoli De Oliveira et al., 2020). در این بررسی با ادغام ویژگی‌های عمیق و اطلاعات معنایی (ویژگی‌های معنایی با سان (فهم صحنه)[۳۷] و یولو[۳۸] استخراج شده است) به‌دست‌آمده از تصاویر، طبقه‌بندی تصاویر و تحلیل احساسات بهبود یافته است. در این پژوهش از یادگیری عمیق و چهار روش شبکه‌های عصبی پیچشی استفاده شده است.

مجموعه‌ی داده‌های این پژوهش تصویری هستند. تصاویر فضای شهری با درنظرگرفتن موقعیت آنها از شبکه اجتماعی فلیکر، استخراج شده و در مرحله آماده سازی و برچسب‌زنی از ۳۰ داوطلب استفاده شده و داده‌ها برچسب‌گذاری شده‌اند (Borth et al., 2013; Chen et al., 2014; Quanzeng et al., 2015 به نقل از Bonasoli de Oliveira et al., 2020). نوآوری این پژوهش، بررسی تأثیر ادغام ویژگی‌های عمیق با ویژگی‌های معنایی است. در واقع در ابتدا در رابطه با صحنه‌ها از هر دو ابزار استفاده شده و ویژگی‌های مستخرج شده از هر دوی آنها به‌عنوان دسته‌بندی برای توصیف صحنه‌ها و اشیا موجود در تصاویر استفاده شده است. نتایج نشان داده که استفاده از ویژگی‌های معنایی، عملکرد تمامی روش‌های شبکه‌های عصبی پیچیده را ارتقا داده است. همچنین براساس نتایج، امکان ویرایش تصاویر، بر عملکرد الگوریتم تأثیر دارد، برای مثال استفاده از تصاویر نشر داده شده در اینستاگرام، برخلاف تصاویر فلیکر، به دلیل آنکه امکان فیلتر، پیام‌های نوشتاری و... را فراهم می‌آورد، مناسب‌تر است.

مقایسه‌ی تجارب تحلیل احساسات شهری توسط شبکه‌های اجتماعی با بهره‌گیری از روش‌های لغوی و مدل‌گرا، در جدول ۱ آورده شده است.

چارچوب مفهومی پژوهش- ارائه‌ی مدل تحلیل احساسات شهری در شبکه‌های اجتماعی مبتنی بر یادگیری ماشین

ازآنجاکه هدف این نوشتار، ارائه‌ی مدل تحلیل احساسات شهری مبتنی بر روش‌های مدل‌گرا است، در ادامه این مدل به‌صورت مبسوط شرح داده می‌شود.

گام اول، در روش یادگیری ماشین، جمع‌آوری داده است. با توجه به نوع داده‌هایی که می‌شود با آنها به تحلیل احساسات پرداخت، داده‌ها در دسته‌های متنی، غیرمتنی و چندوجهی قابل بررسی هستند. داده‌ها را می‌توان از انواع شبکه‌های اجتماعی چون توئیتر، فیس‌بوک، اینستاگرام، فلیکر و... جمع‌آوری نمود. همانطور که پژوهش‌ها نشان داده‌اند، داده‌های متنی عموماً از توئیتر بارگیری شده‌اند، زیرا این شبکه اجتماعی قابلیت‌های مناسبی برای بارگیری اطلاعات دارد. در این بخش در واقع پست‌ها به همراه فراداده خود شامل مکان، زمان، تعداد کامنت‌ها و محتوای پست قابل‌دانلود هستند. نکته‌ای که در این بخش حائز اهمیت است انتخاب زبان در داده‌های

متنی به‌عنوان مجموعه‌های داده‌های ورودی به مدل است. براساس پژوهش‌های انجام شده به زبان فارسی، محدودیت‌هایی در توسعه فرهنگ لغات برای برچسب‌زنی و یا روش‌های پیش‌پردازش متن وجود دارد. در این بخش لازم است تا با توجه به مسئله‌ی پژوهش، موقعیت جغرافیایی و بازه زمانی برای دانلود پست‌ها نیز در نظر گرفته شود زیرا ممکن است در بازه‌های زمانی کوتاه، پست‌های تکراری زیادی وجود داشته باشد. در عین حال تعداد پست‌های موردنیاز برای تحلیل نیز با توجه به مسئله‌ی پژوهش بسیار مهم است زیرا بر دقت و صحت مدل پیش‌بینی‌کننده تأثیر می‌گذارد. در این میان چنانچه

Table 1. Comparison of the experiences of urban sentiment analysis by social networks using lexical and model-oriented methods

Location	Scale	Type and Volume of Data and how Data are Collected	Purpose	Different from this research
Birmingham, UK, 2014	Urban Green Space	Text- Twitter Using the abundance of 10 days. Tool for extracting tweets: Rest API.	Comparison of the effect of labeling methods on classification results	Application and comparison of three different tweet tagging methods: manual, automatic and semi-automatic and identifying the best of them.
Chicago and New York, USA, 2015-2019	City Park (High Line Park/ Park 606)	Data - Twitter The number of Park 606's tweets: 12952 tweets The number of tweets of High Line Park: 165347 tweets Using the VADER. Dictionary, which not only examines the polarity, but also examines the intensity of the feeling.	Sentiment analysis, Identification of user activities and their satisfaction.	-
Boston, USA, 2018	Urban space	Image – Instagram 10 thousand data which include criteria such as date, post content, number of comments, location, latitude and longitude. Using Picodash to extract Instagram data. Using tone analyzer to analyze emotions (Ntshangase, 2018) Presenting data in Geo JSON format	Sentiment classification and providing emotional maps and emotional changes in urban space	Data visualization based on their geographic information. Providing two maps, emojiscape and Trendscape
2014	City Scale	Text- Twitter Twitter about the New York area during Fashion Week on February 6-13, 2014; and in the Boston area; the data of Marathon event on April 15, 2013 were used. The metadata of tweets include: location, time, user ID, post ID, etc. Using classification algorithms and GIS.	Using the methods of computational linguistics and GIS science and an interdisciplinary algorithm for extracting semantic-spatial emotions. Presenting the sentiment analysis process as follows: 1-Filtering 2-Margin writing 3-Seed selection 4a-Sentiment Classification 4b-Tagging tweets 5-Evaluation	Sentiment classification by semi-supervised method and their spatio-temporal identification. Integration of emotions by sensors, Volunteered geographic information and GIS.
Chicago, USA, 2020	Urban open space	Image- Twitter and Flickr. Type of learning: deep learning Data type: Twitter data from DeepSent (includes 1269 Twitter images)/ Outdoor image data from OutdoorSent dataset (includes 40,516 public images available on Flickr related to Chicago) Using several neural network algorithms Using Sun and Yolo to analyze the scene and objects classification	Investigating the effect of combining deep features with Sun and Yolo features	using 4 experimental dataset and 5 convolutional neural network architectures and comparing them, including GG16, Rosnet 50, initial module version 3, DenseNet 169 and the network proposed by You et al. Combining features of Sun (scene understanding and object division) and Yolo (feature of global scenes) with deep features.
USA, 2015	Two sets of 50 cities: Declining and growing cities	Text - Twitter 309 thousand tweets in 2 months. Download tweets including user ID, username, text, longitude, latitude, language and location. Sentiment analysis based on AFINN dictionary. Using Binary Logistic Regression. Ps_match_multi SAS macro Optimization Algorithm.	Investigating the difference between users' feelings of declining and growing cities. Analysis following four criteria: the ratio of positive to negative scores, the ratio of positive to negative tweets, the percentage of positive and negative tweets.	Using the propensity score matching process to determine samples (100 cities). Using the Urban Attitude Program[39]
London, 2015	Urban neighbourhood	Text - Using the SentiHood data set, which includes 5215 sentences related to places. Identify a list of keywords including price, safety, shipping location and general for classification. The use of BRAT's annotation tool and choosing three annotators without expertise. Using the short-term long-term memory algorithm	Sentiment analysis in localities based on LSTM algorithm	Feature-based sentiment analysis using SentiHood datasets

داده‌های دانلود شده متنی کم باشند، روش‌هایی برای تقویت داده‌های متنی وجود دارد. نکته مهم آنکه این روش‌ها صرفاً باید بر روی داده‌های یادگیری اعمال شوند و نه داده‌های تست. داده‌های تست باید بدون تغییر در تعداد حفظ شوند (Pourmostafa Roshan Sharemi, 2017). سه روش برای تقویت داده‌ها وجود دارد: تقویت داده‌ها از طریق داده‌های اضافه؛ تقویت داده‌ها از طریق ترجمه جملات (Pourmostafa Roshan Sharemi, 2017 به نقل از Fadaee et al., 2017)؛ و تقویت داده از طریق جایگزینی با مترادف آنها.

گام دوم، مشتمل بر آماده سازی داده‌ها است که در بحث متن‌کاوی با عنوان پیش‌پردازش متون بیان می‌شود.

گام سوم، مهندسی ویژگی در یادگیری ماشین است که در روش تحلیل احساسات، به دلیل نوع مسئله، ویژگی‌ها در واقع احساسات هستند و بنابراین این بخش مبتنی بر تحلیل معنایی متون و برچسب‌زنی داده‌ها است. سه نوع روش برچسب‌زنی وجود دارد:

۱. روش دستی[۴۰]: حاشیه‌نویسی[۴۱] دستی مستلزم آن است که متن توسط حاشیه‌نویسان انسانی در دسته‌ی احساسات طبقه‌بندی شوند (Jansen et al., 2009).

۲. روش خودکار و اتوماتیک: رویکردهای مختلف برای حاشیه‌نویسی کاملا خودکار وجود دارد، اما روش‌ها معمولاً برای تحلیل احساسات یک متن متکی بر ان-گرام[۴۲] هستند (Barbosa & Feng, 2010 به نقل از Roberts et al., 2018).

۳. نیمه‌خودکار: این روش مبتنی بر حاشیه‌نویسی نیمه نظارت شده در یادگیری ماشین است (Go et al., 2009 ; Pak & Paroubek, 2010 به نقل از Roberts et al., 2018).

گام چهارم، آموزش و تست ماشین است. پس از آماده سازی داده‌ها، داده‌ها برای آموزش ماشین مورد استفاده قرار می‌گیرند عموماً. از میان داده‌های جمع‌آوری شده، ۲۰ تا ۳۰ درصد برای تست استفاده می‌شوند و بقیه برای آموزش ماشین مورد استفاده قرار می‌گیرند. الگوریتم‌های مورد استفاده برای تحلیل احساسات متون که‌عموماً در پژوهش‌ها مورد استفاده قرار گرفته‌اند شامل رگرسیون لجستیک، ماشین بردار پشتیبان، بیز ساده و جنگل تصادفی هستند.

گام پنجم، ارزیابی دقت و صحت مدل است که از طریق ماتریس درهم‌ریختگی[۴۳] انجام می‌شود و چهار شاخص صحت[۴۴]، فراخوانی[۴۵]، دقت[۴۶]، فرمول افیک در آن بررسی می‌شوند (Grandini, Bagli, 2020, 5).

مجموعه داده‌ها و روش‌شناسی پژوهش

این پژوهش از رویکرد یادگیری ماشین نظارت شده بهره برده و به تحلیل احساسات توئیت‌های مربوط به کلان شهرهای کشور پرداخته است. مجموعه داده‌های این پژوهش مشتمل بر ۹۳۰ توئیت است که در بازه زمانی ۱۰ سال از ۱۳۹۱ تا ۱۴۰۰ جمع‌آوری شده‌اند. نکته مهم آن است که توئیت‌های اولیه بیش از ۵۰۰۰ توئیت به همراه متن و تصویر بوده است که در فرایند برچسب‌زنی با توجه به تأکید بر توئیت‌های مربوط به شهر و فضا در شهر، ۸۰ درصد آنها به دلیل عدم ارتباط آنها به شهر حذف شده است. عبارت جستجو شده مشتمل بر نام شهر و محدوده‌های گردشگری بوده تا تعادلی بین داده‌های مثبت و منفی برقرار شود. توئیت‌ها از طریق توئیتراستریمینگ بارگیری شده و فرا داده همراه با متن مشتمل بر تعداد ری‌توییت‌ها، تعداد پسندها، توئیت‌آی.دی. عبارت جستجو، زبان و موقعیت بوده است. مجموعه داده‌ها پس از مراحل استاندارد و نرمال سازی جهت آموزش ماشین مورد استفاده قرار گرفته‌اند. حجم داده آموزش و تست در این پژوهش ۸۰ به ۲۰ بوده است. با توجه به رویکرد نظارت شده، داده‌ها توسط محقق با سه برچسب منفی، خنثی و مثبت برچسب‌زنی شده‌اند و در مواردی که محقق در برچسب‌زنی دچار تردید بوده است، از نظر دو متخصص دیگر بهره برده شده است. به‌صورت کلی هم از یادگیری ماشین و هم یادگیری عمیق بهره برده شده است. به منظور بررسی اعتبار مدل و تست آن، از ماتریس در هم آمیختگی استفاده شده است.

جدول ۲، نمونه‌ای از جدول داده‌های بارگیری شده از توئیتر در قالب فایل سی.اس.وی. را نشان می‌دهد.

Table 2. An example of the data downloaded from Twitter in CSV file format

Search term	#City-Isfahan
Tweet ID	1534803796555091970
Number of likes	42
Number of retweets	41
Position	Iran
Language	Fa.
Text	Today, a group of farmers of Zarin City #Isfahan protested against the closure of Zayandeh River https://t.co/OyzMydN8xd

تحلیل و یافته‌ها

در این پژوهش، کدها با استفاده از پایتون و در فضای ابری گوگل کولب نوشته شده است. کتابخانه‌های موردنیاز برای یادگیری ماشین شامل پانداز، کراس و سایکیت لرن[۴۷] بوده و در ابتدا فراخوانی شده‌اند. کتابخانه پانداز در بررسی متن کاربرد مناسبی دارد و کتابخانه کراس به دلیل انتخاب رویکرد یادگیری عمیق و شبکه عصبی بهره مورد توجه قرار گرفته است.

گام اول- فراخوانی داده‌ها: در ابتدا لازم است که داده‌ها فراخوانی شوند و مشخص شود که تعداد

داده‌های گم شده در هر ویژگی به چه میزان است. در این بررسی مشخص شد که ۳۷۳ سلول مربوط به موقعیت خالی است.

گام دوم- پیش‌پردازش داده‌ها: در این بخش لازم است که نسبت به داده‌های گمشده اقدام شود. دو نوع داده در فرایند پیش‌پردازش تشخیص داده شد، داده‌های integer وfloat و داده‌های string یا object. اگر نوع داده از نوع دوم باشد، می‌توان از روش بیشترین تکرار[۴۸] یا مد استفاده نمود. برای مثال چنانچه موقعیت مشخص نباشد، می‌توان از روش بیشترین تکرار بهره برد و سلول‌های خالی را پر کرد. بررسی داده‌ها همچنین نشان داد که چند ویژگی رشته‌ای در داده‌ها وجود دارند که باید به کد تبدیل شوند. برای مثال داده‌های مربوط به موقعیت و عبارت جستجو. با بهره‌گیری از روش Label Encoder برای هر رشته یک کد اختصاص داده شد و داده‌ها جهت آموزش آماده گردید.

گام سوم- نرمال سازی داده‌ها: برای نرمال سازی داده‌ها از روش‌های مختلفی استفاده می‌شود که شامل MinMaxScaler و StandardScaler هستند. در این پژوهش از روش اول استفاده شده است. هدف این بخش نرمال کردن داده‌ها بین صفر و یک است. فرمول ریاضی روش مورد استفاده در این پژوهش، به شرح زیر است:

$$\text{X scaled} = \frac{X - Xmin}{\text{Xmin} - \text{Xmax}} \quad \textbf{(1)}$$

گام چهارم: تعیین حجم داده‌های تست و آموزش

گام پنجم: آموزش و تست ماشین

مرحله اول: تمیزکردن متون

برای داده‌های متنی لازم است که داده‌ها، تمیز[۴۹] شوند. با توجه به اینکه زبان متون در این پژوهش فارسی بوده از کتابخانه هضم[۵۰] استفاده شده است. این کتابخانه برای انجام پردازش‌های لازم بر روی زبان فارسی توسط دانشجویان دانشگاه علم و صنعت ایران در سال ۱۳۹۲ به‌صورت متن باز و با استفاده از کتابخانه ان. ل.تی.کی. منتشر شده است. از قابلیت‌های هضم تمیز و مرتب‌کردن متن، جداسازی جمله‌ها و واژه‌ها، ریشه‌یابی، تحلیل صرفی جمله، تجزیه نحوی جمله و ... است[۵۱].

برای تمیزکردن داده‌های متنی موارد زیر موردتوجه قرار گرفته است:

- با توجه به اینکه حروف الفبا مدنظر است، تمام اعداد و علائم از متن حذف شد.
- حذف لینک‌ها انجام شد.
- فضاهای اضافی ابتدا و انتهای حذف شدند.
- نرمال سازی داده‌ها در زبان فارسی با دستور normalize Hazm انجام شد.
- توکنایز کردن کلمات مدنظر قرار گرفت.
- ریشه‌یابی کلمات انجام شد.
- ایست واژه‌ها حذف شدند.

در جدول ۳ نمونه متن اولیه و متن تمیز شده ارائه شده است.

Table 3. Sample of original text and cleaned text

Sample text
The Milad Tower is also lost in this #air_pollution https://t.co/j2smPUTBNG
The Milad Tower was also lost again in this air pollution

مرحله دوم: توکنایز کردن

با استفاده از توکنایزر کراس، امکان شماره گذاری جملات و تبدیل آنها به بردار وجود دارد. اندازه فرهنگ لغت نیز با تست اعداد مختلف به دست آمد و مقداری عددی اختصاص یافته به MAX_NB_WORDS، ۲۰۰ هزار کلمه در نظر گرفته شد.

مرحله سوم: همسان سازی طول جملات

از آنجا که در فرایند توکنایز کردن جملات، طول جملات یکسان نیستند، لازم است تا با استفاده از دستور Zero Padding، آنها را یکسان نمود.

شکل ۱، نمونه ای از اعمال دستور زیرو پدنیگ برای یکسان‌سازی طول جملات را نشان می‌دهد.

مرحله چهارم: تعبیه‌سازی کلمات[۵۲]

در ارائه‌ی داده‌های متنی به ماشین به صورت قابل فهم، می‌توان از روش‌های متعددی استفاده نمود، در این پژوهش بِه‌جای استفاده کردن از روش برداری کردن که‌صرفاً جملات را به بردار تبدیل می‌کند و اطلاعات مفیدی دیگری در اختیار قرار نمی‌دهد، از روش تعبیه‌سازی کلمات استفاده شده است. در واقع داده‌های متنی، ساختاری ترتیبی دارند و این روش بهترین روش برای نمایش داده‌های غیرعددی است که ساختاری ترتیبی دارند. در این رویکرد داده‌هایی با ابعاد کمتر و اطلاعات بیشتر

```
[ 395  278  248  115    2 1042   70  278  157  278 1444 2242 1445  798
  395 2243 1446 2244 2245  189  105  210  106  132 2246  278  248  798
 1447  189 2247  657 1448    1  658    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0
                                  0    0]
```

Fig. 1. An example of using the zero padding command to equalize the length of sentences

ایجاد می‌شود و هدف آن نگاشت معنا به یک فضای هندسی است. درعین‌حال در کتابخانه کراس نیز دو مدل اصلی ترتیبی و فانکشنال پشتیبانی می‌شوند.

از میان الگوریتم‌هایی چون ورد.وک.۲، گلاو و فست تکست که برای تعبیه‌سازی کلمات مورد استفاده قرار می‌گیرند؛ در این پژوهش از فست تکست[۵۱] استفاده شده است. این کتابخانه توسط آزمایشگاه تحقیقات هوش مصنوعی فیس‌بوک ایجاد شده است. این مدل این امکان را به پژوهشگر می‌دهد که یک الگوریتم یادگیری بدون نظارت یا یادگیری نظارت شده برای به‌دست‌آوردن نمایش‌های برداری به کلمات را ایجاد کند. فیس‌بوک مدل‌های از پیش آموزش دیده را برای ۲۹۴ زبان در دسترس قرار داده است.

بدین ترتیب در این بخش وزن‌های موردنیاز برای یادگیری عمیق مشخص می‌شود و این امر نه به‌صورت اتفاقی بلکه با بهره گیری از کتابخانه فست تکست انجام شده است. طول لایه embedding در این پژوهش ۳۰۰ در نظر گرفته شده است.

مرحله پنجم: آموزش ماشین

در این بخش برای آموزش ماشین هم از الگوریتم‌های مورد استفاده در یادگیری ماشین و هم یادگیری عمیق استفاده شد.

الف) آموزش ماشین با الگوریتم‌های رگرسیون لجستیک، ماشین بردار پشتیبان و جنگل تصادفی

در ابتدا آموزش ماشین براساس ۳ الگوریتم که در بسیاری از پژوهش‌های مربوط به تحلیل احساسات متن مورد استفاده قرار می‌گرفتند (Saeidi et al., 2016)، آموزش داد ه شد. براساس نتایج تست که براساس ماتریس در هم آمیختگی ارائه شده، دقت ماشین آموزش دیده در تعیین قطبیت متن در سه قطبیت منفی، خنثی و مثبت به‌صورت زیر است. از میان سه الگوریتم استفاده شده، ماشین بردار پشتیبان و جنگل تصادفی از الگوریتم دیگر بهتر عمل کرده‌اند.

با توجه به اینکه دقت مدل در بهترین حالت حدود ۷۰ درصد بوده؛ از یادگیری عمیق نیز برای آموزش ماشین استفاده شد تا مشخص شود آیا نتیجه بهتری براساس آن حاصل می‌شود یا خیر.

در جدول ۴ دقت ماشین در تعیین قطبیت متن در الگوریتم‌های مورد استفاده، آورده شده است.

Table 4. Accuracy of the machine in determining the polarity of the text in the used algorithms

Algorithm	Negative polarity	Neutral polarity	Positive polarity
Logistic Regression	68	51	63
Support Vector Machine	73	57	70
Random Forest	71	56	68

ب) آموزش ماشین با الگوریتم شبکه عصبی کانولوشنی و الگوریتم هیبریدی

در ابتدا با استفاده از شبکه عصبی کانولوشنی ماشین آموزش داده شد، نتایج دقت مدل نشان داد که مدل تا ۷۵ درصد قابلیت پیش‌بینی پذیری دارد.

در ادامه تلاش شد با نوشتن یک الگوریتم هیبریدی مبتنی بر شبکه عصبی کانولوشنی، دقت پیش‌بینی‌پذیری مدل ارتقا یابد. معماری این شبکه به این صورت است که دو نوع داده به‌عنوان ورودی به شبکه عصبی در نظر گرفته شده، داده‌های متن و همچنین سایر ویژگی‌های موجود در مجموعه داده‌ها شامل موقعیت، ری‌توییت‌ها، تعداد پسندها، کد شهرها و محتوای جستجو شده (به‌عنوان متادیتا). بنابراین براساس این ورودی و خروجی (طبقه‌بندی براساس قطبیت متن که توسط محقق انجام شده است) ماشین آموزش دید و در نهایت تست شد.

همانطور که در معماری الگوریتم هیبریدی مشاهده می‌شود (شکل ۲)، اهمیت متن معادل ۹۰ درصد و اهمیت متادیتا ۱۰ درصد لحاظ شده است. لازم به‌ذکر است که درصدهای متفاوتی به اهمیت هر یک از ورودی‌ها داده شد و دقت پیش‌بینی‌پذیری مدل بررسی شد. همانطور که نتایج تست مدل نشان می‌دهد، الگوریتم طراحی شده تا ۴ درصد قابلیت پیش‌بینی‌پذیری ماشین را ارتقا داده است.

مرحله ششم: اعتبارسنجی و تست ماشین

براساس ماتریس درهم‌آمیختگی مشاهده می‌شود که میزان دقت‌ماشین با بهره گیری از الگوریتم هیبریدی در تعیین قطبیت متن حدود ۸۰ درصد است و شبکه طراحی شده مناسب عمل می‌کند.

شکل ۳ نشان می‌دهد که در تست مدل در طبقه صفر که در این پژوهش معادل قطبیت منفی است و در طبقه دو که نشان دهنده قطبیت مثبت است الگوریتم استفاده شده در شناسایی قطبیت متن، حدود ۸۰ درصد دقت داشته است. در رابطه با طبقه خنثی الگوریتم مورد استفاده به دلیل تعداد کم داده‌های طبقه خنثی حدود ۶۰ درصد دقت داشته است.

نتیجه‌گیری

در این مقاله، تحلیل احساسات مبتنی بر روش‌های مدل‌گرا - یادگیری ماشین و یادگیری عمیق - مورد تأکید قرار گرفت و ازاین‌رو ضمن مقایسه آن با روش‌های سنتی و روش‌های لغوی، فرایند تحلیل احساسات شهری تدوین شد و سطوح مختلف فرایند به‌تفصیل شرح داده شد. بررسی نمونه‌های جهانی و تجربه تحلیل احساسات شهری با شبکه‌های اجتماعی در مقیاس‌های مختلف و همچنین مکان نگاشت تحلیل‌های احساسی نیز بررسی شد که نتایج آنها نشان داد که روش مدل‌گرا جایگاه مناسب و تکمیلی در کنار سایر روش‌های تحلیل احساسات

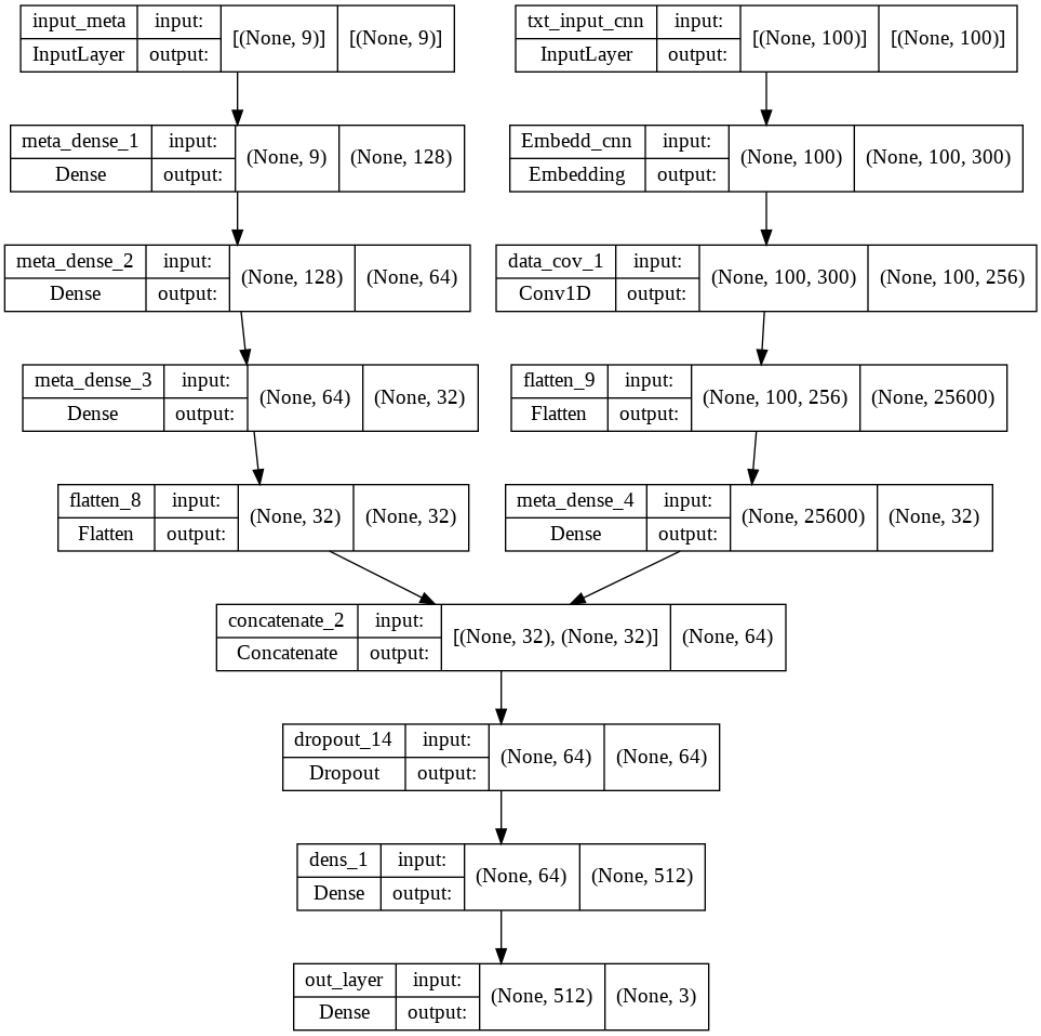

Fig. 2. Designed hybrid network architecture

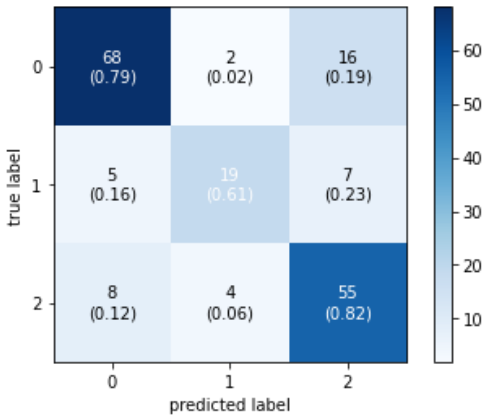

Fig. 3. Confusion matrix in the designed neural network algorithm

داشته و می‌توانند در مقیاس‌های کلان به‌عنوان روشی مناسب برای تحلیل احساسات کاربران از یک سو و پیش‌بینی رفتار آنها در شهر باشد.

همانطور که بیان شد این روش‌ها، دارای مزایای متعددی هستند و در مسیر پروژه‌های شهری می‌توانند برای تحلیل وضع موجود یا پیش‌بینی وضعیت مفید فایده باشند. ضمن آنکه نسبت به روش‌های سنتی، کم‌هزینه‌تر بوده، سرعت بیشتری داشته و از دقت کافی برخوردار هستند.

همچنین در این پژوهش کاربست روش مدل‌گرا در تحلیل احساسات مربوط به کلان شهرها در کشور براساس داده‌های فضای مجازی نیز انجام شد. با توجه به قابلیت مناسب ماشین آموزش داده شده، این ماشین قابلیت پیش‌بینی قطبیت داده‌های شهری را دارد. به این معنی که با استفاده از داده‌های متنی که در شبکه‌های اجتماعی منتشر می‌شوند، می‌توان به تحلیل احساسات کاربران پرداخت. بدون شک چنانچه این داده‌ها ژئوبیس باشند، قابلیت مکانی کردن احساسات نیز وجود دارد و از این طریق با سرعت، دقت مناسب و هزینه کم، می‌توان به ممیزی کلی فضاها در شهر پرداخت و از این طریق با شناسایی فضاهایی که حس منفی به کاربران منتقل می‌کنند، به بررسی دلایل و رفع آنها اقدام نمود. این پژوهش از دو حیث واجد نوآوری بوده است، ۱) تهیه مجموعه داده‌های مربوط به احساسات کاربران فارسی‌زبان مربوط به شهر و ۲) تحلیل احساسات شهری در کشور با استفاده از یادگیری ماشین در حوزه برنامه‌ریزی و طراحی شهری. برخی از محدودیت‌های این پژوهش شامل: محدودیت دسترسی به تمامی داده‌های نشر داده شده در توئیتر با استفاده از توئیتر استریمینگ؛ حجم کم داده‌های در دسترس و مرتبط؛ استفاده کم کاربران از توئیتر به دلیل فیلترینگ؛ و در دسترس نبودن منابع مالی برای تهیه مجموعه بزرگتر داده‌ها بوده است.

پی‌نوشت

1. E.G.Boring
2. Boring
3. Exciting
4. Scary
5. Consoling
6. Depressing
7. Uplifting
8. Dull
9. Depressing
10. Open
11. Inspiring
12. Inspiring
13. Impressive
14. Peaceful
15. Yelp & Tripadvisor
16. Flickr
17. Tumblr
18. Pang
19. Naïve Bayes (NB), Maximum Entropy & Support Vector Machine (SVM)
20. Lexical method
21. Stanford University
22. The science of getting computers to act without being explicitly programmed
23. Judith Hurwitz
24. Daniel Kirsch
25. Apirest
26. Fully Automated Annotation
27. ANEW Affective Words English for Norms
28. Valence score
29. Line High
30. VADER
31. GeoNames
32. Aspect-based
33. LSTM
34. AFINN
35. Nielsen Årup Finn

۳۶. با استناد به پژوهش هلند و رنسکی (۲۰۱۵)، در چندین مطالعه تحقیقاتی از این فرهنگ لغت استفاده شده است، از جمله تحلیل توییت‌های منتشر شده از نیو بدفورد، ام.ای. بین ۹ فوریه ۲۰۱۴ و ۳ آوریل ۲۰۱۴ (Hollander et al., 2014)، شناسایی نظرات ضد واکسن از توییت‌ها (Brooks, 2014)، ارزیابی بیش از ۵۰۰۰ آگهی در مجلات تجاری (Abrahams et al., 2013) و به عنوان بخشی از مدل پیش‌بینی نوسانات در بازارهای جهانی ارز (Jin et al., 2013).

37. SUN (Scene UNderstanding)
38. YOLO (You Only Look Once)

۳۹. آزمایشگاه نگرش‌های شهری، داده‌های بزرگ و ابعاد روانشناختی برنامه‌ریزی شهری و سیاستگذاری عمومی را مطالعه می‌کند. این آزمایشگاه از طریق استفاده از ابزارهای پیچیده جمع‌آوری داده، تجزیه و تحلیل احساسات، فناوری ردیابی چشم و ابزارهای روانشناختی مرسوم، دانش را پیش می‌برد و ابزارهای جدیدی را برای بهبود برنامه‌ریزی و سیاست توسعه می‌دهد (Urban Attitudes Lab, 2 May 2022).

40. Manual Annotation
41. Annotation
42. N-gram
43. Confusion Matrix
44. Precision
45. Recall
46. Accuracy
47. Pandas, Keras, Sickit-learn
48. Most frequency
49. Clean
50. Hazm

۵۱. وب سایت خانه بیگ دیتای ایران (۱۴۰۰). پردازش متن تا Jhazm نسخه جاوا کتابخانه هضم برای پردازش زبان فارسی [نوشته وب سایت]. بازیابی شده در ۱۰ اسفند ۱۴۰۰ از:
https://bigdata-ir.com/%d9%be%d8%b1%d8%af%d8%a7%d8%b2%d8%b4-%d9%85%d8%aa%d9%86-%d8%a8%d8%a7-jhazm/

52. Embedding
53. Fasttext

تشکر و قدردانی

موردی توسط نویسنده گزارش نشده است.

تعارض منافع

نویسنده اعلام می‌دارد که در انجام این پژوهش هیچ‌گونه تعارض منافعی برای ایشان وجود نداشته است.

تاییدیه‌های اخلاقی

نویسنده متعهد می‌شود که کلیهٔ اصول اخلاقی انتشار اثر علمی را براساس اصول اخلاقی COPE رعایت کرده و در صورت احراز هر یک از موارد تخطی از اصول اخلاقی، حتی پس از انتشار مقاله، حق حذف مقاله و پیگیری مورد را به مجله می‌دهد.

منابع مالی/ حمایت‌ها

موردی توسط نویسنده گزارش نشده است.

مشارکت و مسئولیت نویسندگان

نویسنده اعلام می‌دارد به‌طور مستقیم در مراحل انجام پژوهش و نگارش مقاله مشارکت فعال داشته و به طور برابر مسئولیت تمام محتویات و مطالب گفته‌شده در مقاله را می‌پذیرد.

References

1. Ayodele, T. (2010). *Types of Machine Learning Algorithms*, University of Portsmouth, United Kingdom, 21-48.
2. Bonasoli de Oliveira, W. et al. (2020). OutdooeSent: Setiment analysis of urban outdoor images by using semantic and deep features. *ACM transactiona on information systems*, 1(1), 1-29.
3. Breiman, L. (2001). *Random Forests*, Berkeley, University of California, 1-33.
4. Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *22nd acm sigkdd international conference on knowledge discovery and data mining,* America, 857-794.
5. Devika, M. D., Sunitha, C. and Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, Vol 87.
6. EURISY, 2017 (https://goodcitylife.org/)
7. Fathullah, A., & S.Willis, K. (2018). Engaging the Senses: The Potential of Emotional Data for Participation in Urban Planning. *Urban Science*, 1-21. doi:10.3390/urbansci2040098
8. Fooladi, S., Farsi, H., Mohamadzadeh, S. (2019). Detection and classification of skin cancer using deep learning. *Journal of Birjand University Medical Science*, 26 (1), 44-53
9. Gehl, J. (2010). Cities for People. *(Translated*

by Ali Ghafari & Leila Ghafari). Tehran: Elm-e Memar Publication. [In Persian]
10. Grandini, M. Bagli, E. Visani, G. (2020). Metrics for Multi-Class Classification, An Overview, *Universita degli Studi di Bologna*, Italy, 1-17.
11. Grutter, J. K. (1987). *Asthetik der Architektur: Grundlagen der Architektur-Wahrnehmung*, (Trans. Jahanshah Pakzad & Abdolreza Homayoon). [In Persian]
12. Hastie, T. Rosset, S. Zou, H. Zhu, J. (2009). Multi-Class Adaboost. *Statistics and its Interface*, VOL .2, 349-360.
13. Hogertz, C. (2010). Emotions of the urban pedestrian: sensory mapping. *Pedestrians' Quality Needs*, 31-52.
14. Hollander, J. B. & Renski, H. (2015). *Measuring urban attitudes using Twitter: an exploratory study.* Lincoln Institute of Land Policy.
15. Houtkamp, J. M. (2012). *Affective appraisal of virtual environments*. Leiden University.
16. Huang, H., Gartner, G., Klettner, S., & Schmidt, M. (2014). Considering affective responses towards environments for enhancing location based services. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 93-96. doi:10.5194/isprsarchives-XL-4-93-2014
17. Hurwitz, J. Kirsch, D. (2018). *Machine Learning For Dummies*, IBM Limited Edition, New Jersey.
18. Hutto, C.J., Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, May 2014. Available online: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109 (accessed on 6 August 2019)
19. Jahid, Z., Nasiri, J. (2016). *Sentiment analysis using sentiment measurement. The third International Conference on Information Technology*. Computer Engineering and Telecommunications. September. [In Persian]
20. Ji et al., (2015). *Cross-Modality Sentiment Analysis for Social Multimedia*. IEEE International Conference on Multimedia Big Data
21. Kaplan, A.M. & Haenlein, M. (2010). Users of theWorld, Unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. http://dx.doi.org/10.1016/j.bushor.2009.09.003
22. Keith, d (2019). A Brief History of Machine learning, Dataversity site, https://medium.com/geekculture/decision-tree-random-forest-explained-with-project-c2c8ef1597ee.
23. Kim, E. et al. (2018). Urban emotion: The interrogation of social media and its implication within urban context. *The cognitive city*, 2, 475-482.
24. Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79, 1–14.
25. Lin, Y.; Geertman, S. (2019). Can Social Media Play a Role in Urban Planning? A Literature Review. In (Eds. Geertman, S., Zhan, Q., Allan, A., Pettit, C.) *Computational Urban Planning and Management for Smart Cities*; Springer International Publishing: Cham, Switzerland, 69–84.
26. Liu, B., (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. s.l.: Cambridge University Press.
27. Malhotra, V. & Semwal, M. (2019). *Comparison of 3 Supervised Machine Learning Models*, Thesis of Phd, University of Windsor .
28. Mousavi S M, Ebadi H, Kiani A. (2019). Provide an Optimal Deep-network Method for Spectral-spatial Classifying of High Resolution Images. *Journal of Geomatics Science and Technology*, 9 (2), 151-170. URL: http://jgst.issgeac.ir/article-1-814-fa.html. [In Persian]
29. Osborne, T., & Jones, P. (2017). Biosensing and geography: a mixed methods approach. *Applied Geography*, 1-19.
30. Pakzad, J. (2006). *Theoretical foundations and process of urban design*. Tehran: Shahidi Publication. [In Persian]
31. Poria, Soujanya. Et al. (2017). Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing*, 261, 217-230.
32. Pourmostafa Roshan Sharemi, J., Abbasi Sarabestani, P., Mirroshandel, S. A. (2017). Presenting a sentiment analysis system in Persian using deep learning models. *Journal of Language and Linguistics*, 1-15. [In Persian]
33. Pykett, J., Chrisinger, B. W., Kyriakou, K., Osborne, T., Resch, B., Stathi, A., & Whittaker, A. C. (2020). Urban Emotion Sensing Beyond 'Affective Capture': Advancing Critical Interdisciplinary Methods. *International Journal of Environmental Research and Public Health*, 1-22. doi:10.3390/ijerph17239003
34. Rad, R., Musavi, M., & Vardi, F. (2021). Plant Disease Detection by Image Processing of Their Leaves Based on Deep-Learning. *Agriultural Mechanization and Systems Research*. 76(76), 49-68. [In Persian]
35. Rastgoo, R. & Kiani, K. (2019). Face recognition using fine-tuning of Deep Convolutional Neural Network and transfer learning. *Journal of Modeling in Engineering*, 17(58), 103-111. [In Persian]
36. Resch, B. et al. (2014). Urban emotions-Geo-semantic emotion extraction from technical sensors, human sensors and crowd sourced data. In G. Gartner & H. Huang (Eds.), *Progress in location-based service 2014,* Springer.
37. Roberts, H. et al. (2018). Investigating the emotional responses of individuals to urban green space using Twitter data: a critical comparison of tree different methods of sentiment analysis. *Urban Planning*, 3 (1), 21-33.

38. Saeidi, M. et al. (2016). *SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods.* https://arxiv.org/abs/1610.03771
39. Safavian, R. & Landgrebe, D. (1991). A Survey Of Decision Tree Classifier Methodology, *Reprinted From IEEE Transactions On Systems, Man, and Cybernetics*, Vol. 21, No. 3, 660-674.
40. Sdoukopoulos, A. et al. (2018). Use of social media for assessing sustainable urban mobility indicators. *International Journal of sustainable development and planning*, 13(2), 338-348.
41. Setiawan, Esther Irawati et al. (2021). Multiview sentiment analysis with Image-Text-Concept features of Indonesian social media posts. *International Journal of Intelligent Engineering & Systems*, 14(2), 521-535.
42. Shah Talebi, N., Karegar, M. J., Mirzaei, K. (2015). Examining opinion polling models and analyzing user sentiments in the web environment. *The second International Web Research Conference*, May 8-9. [In Persian]
43. Sim, J. et al. (2020). Tweeting the high line life: a social media lens on urban green spaces. *Sustainablity*, 12, 1-18. SUN397.tar.gz
44. Yousofi Moteghaed, M., Sabohi, H. (2018). *A review of sentiment analysis of social networks in the field of polarity*. International Web Research Conference, May 3- 5, Tehran, Iran. [In Persian]
45. Zhai, C.; Massung, S. (2016). *Text. Data Management and Analysis: A Practical Introduction to Information Retrieval and Text. Mining; Morgan & Claypool: Williston*, ND, USA.
46. Zhang, Y., Jin, R., Zhou, Z.H. (2010). Understanding bag-of-words model: A statistical framework. Int. J. Mach. Learn. *Cyber*. 1, 43–52.