Journal of Business Management
Print ISSN: 2008-5907
Online ISSN: 2423-5091

Customer Reviews Classification with Text Mining and Deep Learning Approach (Case Study: Digikala Customers Reviews)

Parham Parnian
Msc. Student, Department of Computer Engineering, Faculty of Computer Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. E-mail: parnian@eng.uk.ac.ir

Abstract

Objective
Today, people face different decision-making criteria when purchasing products and services. One of these criteria is using the reviews of the previous purchasers of products and services. A large volume of reviews is seen as a challenge for these people. The present study aimed to create a model to yyyyyyy yyyyyyyyyyy ssss and to classify their reviews to solve the mentioned challenge.

Methodology
hhe rr eeett ddddy nngggggggggg gge uuy'''' sssssss ss mbbeee hhnnss uuddddddddd hhe iiii aaaa Website from 2015 to 2016. To analyze the sentiments, and to classify the reviews, deep learning-based algorithms, and convolutional networks, subtypes of deep networks, were suggested. Prior to preprocessing and homogenizing the data, the study used a pre-trained Fastext model to convert the words into integer vectors and deliver them as inputs to the proposed deep network.

Findings
To train the selected model, the training algorithm was carried out on it 90 times. To validate the performance of the selected model, confusion matrix, accuracy, recall, F1-score, and precision rate criteria were used.

Conclusion
The present study used the deep networks approach, convolutional networks, and bidirectional long short-aaaa mmmrr y oorrrr yyyyhhe uuy'''' sssssss ss hhe mbbeeehhnne ooom hhe eeeeeee eee ee tt 93% accuracy, and after 90 training periods.

Keywords: Deep learning, Convolutional neural networks, Reviews classification, Text mining.

Citation: Parnian, Parham (2022). Customer Reviews Classification with Text Mining and Deep Learning Approach (Case Study: Digikala Customers Reviews). *Journal of Business Management*, 14(4), 675-694. https://doi.org/10.22059/JIBM.2022.334338.4255 (*in Persian*)

Journal of Business Management, 2022, Vol. 14, No.4, pp. 675- 694
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/JIBM.2022.334338.4255
Article Type: Research Paper
© Authors
Received: December 06, 2021
Received in revised form: August 01, 2022
Accepted: August 15, 2022
Published online: January 10, 2023

مدیریت بازرگانی
شاپا چاپی: ۵۹۰۷-۲۰۰۸
شاپا الکترونیکی: ۵۰۹۱-۲۴۲۳

طبقه‌بندی نظرهای مشتریان با رویکرد متن‌کاوی و یادگیری عمیق (نمونه موردی: نظرهای کاربران وبسایت دیجی‌کالا)

پرهام پرنیان*
دانشجوی کارشناسی ارشد، بخش مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان، کرمان، ایران. رایانامه: parnian@eng.uk.ac.ir

چکیده

هدف: امروزه افراد هنگام خرید محصول یا تهیه خدمات، معیارهای متفاوتی را برای تصمیم‌گیری در نظر می‌گیرند. یکی از این معیارها، اطلاع از نظر خریداران قبلی محصولات و خدمات است؛ اما حجم زیاد نظرها نیز، چالشی است که پیش روی این افراد قرار دارد. پژوهش حاضر با هدف ایجاد مدلی جهت تحلیل احساسات کاربران و طبقه‌بندی نظر آن‌ها برای حل این چالش اجرا شده است.

روش: پژوهش حاضر روی نظرهای خریداران تلفن همراه از وبسایت دیجی‌کالا، طی سال‌های ۱۳۹۴ تا ۱۳۹۵ انجام شده است. به‌منظور تحلیل احساسات و طبقه‌بندی نظرها، الگوریتم‌های مبتنی بر یادگیری عمیق و شبکه‌های پیچشی (کانوولوشن) که نوعی از شبکه‌های عمیق هستند، پیشنهاد شده است. در این پژوهش پس از پیش‌پردازش داده‌ها و یکسان‌سازی آن‌ها با استفاده از مدل از پیش تعلیم‌دیده فست تکست، کلمه‌ها به بردارهایی از اعداد صحیح تبدیل و به‌عنوان ورودی به شبکه عمیق پیشنهادی تحویل داده شدند.

یافته‌ها: جهت تعلیم مدل منتخب این پژوهش، ۹۰ مرتبه الگوریتم آموزشی روی آن اجرا شد. برای صحت عملکرد مدل منتخب از معیارهای ماتریس گیجی، دقت، بازخوانی، معیار F و میزان دقت استفاده شد.

نتیجه‌گیری: در پژوهش حاضر با رویکرد شبکه‌های عمیق و با استفاده از شبکه‌های پیچشی و حافظه طولانی کوتاه‌مدت دوطرفه، پس از ۹۰ دوره آموزش، توانستیم با دقت ۹۳ درصد عقاید خریداران تلفن همراه در وبسایت دیجی‌کالا را طبقه‌بندی کنیم.

کلیدواژه‌ها: شبکه‌های پیچشی، طبقه‌بندی نظرها، متن‌کاوی، یادگیری عمیق.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

استناد: پرنیان، پرهام (۱۴۰۱). طبقه‌بندی نظرهای مشتریان با رویکرد متن‌کاوی و یادگیری عمیق (نمونه موردی: نظرهای کاربران وبسایت دیجی‌کالا). *مدیریت بازرگانی*، ۱۴(۴)، ۶۷۵- ۶۹۴.

مدیریت بازرگانی، ۱۴۰۱، دوره ۱۴، شماره ۴، صص. ۶۷۵- ۶۹۴
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی
© نویسندگان
تاریخ دریافت: ۱۴۰۰/۰۹/۱۵
تاریخ ویرایش: ۱۴۰۱/۰۵/۱۰
تاریخ پذیرش: ۱۴۰۱/۰۵/۲۴
تاریخ انتشار: ۱۴۰۱/۱۰/۲۰
doi: https://doi.org/10.22059/JIBM.2022.334338.4255

مقدمه

با پیشرفت و توسعهٔ چشمگیر اینترنت در سال‌های اخیر، شرایط مناسبی جهت انجام فعالیت‌های فروشگاه‌های اینترنتی فراهم شده است. همچنین با مشاهدهٔ اطلاعات آماری سال ۲۰۲۱، تعداد کاربرانی که به اینترنت دسترسی دارند، از ۵ میلیارد عبور کرده است که از این تعداد، سهم ایران ۷۸ میلیون کاربر اینترنت است (آمار جهانی اینترنت[1]).

سرعت زیاد گسترش اینترنت، باعث شد که در زندگی روزمرهٔ کاربران، به‌خصوص کاربران ایرانی، مفهومی کلیدی با عنوان تجارت الکترونیک افزوده شود (بخشی‌زاده و حاجی جعفر و نصیری، ۱۳۹۷). وبسایت‌های تجارت الکترونیک جهت خرید، فروش کالا و خدمات استفاده می‌شوند. از طریق این وبسایت‌ها، افراد کالا یا خدمات محبوب خود را می‌بینند، نظر سایر خریداران را بررسی می‌کنند و در نهایت، جهت خرید کالا یا خدمتی که نیاز دارند، تصمیم می‌گیرند (موروگاوالی، بگیرتن، سایپاراسانت و آروبندکومار[2]، ۲۰۱۷). در سال‌های اخیر با رشد شبکه‌های اجتماعی و به‌تبع آن، افزایش محتوای تولید شده در این رسانه‌ها، افراد برای تصمیم‌گیری‌های مهم خود، از جمله خرید محصول یا خدمت و حتی انتخاب‌های سیاسی، از نظر دیگران در این رسانه‌ها بهره می‌برند. اغلب اطلاعاتی که در این شبکه‌ها استفاده می‌شود، به‌شکل متن هستند که با توجه به حجم عظیم داده‌های متنی تحلیل، کشف دانش نهفته در این داده‌ها بسیار ارزشمند خواهد بود (لیو[3]، ۲۰۱۲). تحلیل نظر مشتریان می‌تواند به شناخت ترجیحات و افزایش رضایت مشتریان کمک کند. وبسایت‌های تجارت الکترونیک با تحلیل نظر کاربران خود، می‌توانند به قوت‌ها و ضعف‌های خود پی ببرند. بنابراین تحلیل و پردازش این داده‌های متنی حجیم و ساختارنیافته، جهت توسعه و بهبود تجربهٔ مشتریان، از اهداف مهم و کلیدی کسب‌وکارهای تجارت الکترونیک خواهد بود. دیجی‌کالا[4] که یکی از فروشگاه‌های اینترنتی پیشرو در ایران است، مشتریان خود را تشویق می‌کند تا با ثبت نظرشان دربارهٔ کالاهایی که خریداری کرده‌اند، فرایند تصمیم‌گیری را برای دیگران تسهیل کنند. اما مشکلی که کاربران را درگیر می‌کند، مسئلهٔ کمبود وقت است که باعث می‌شود کاربران نتوانند همهٔ نظرها را بررسی کنند و تنها به بررسی امتیاز سایر کاربران به محصولات اکتفا کنند که این امر گمراه‌کننده‌ای خواهد بود. تحلیل و بررسی نظرهای ثبت‌شده دربارهٔ محصولات، با استفاده از روش‌های یادگیری عمیق، به‌کمک شبکه‌های پیچشی[5] که محور اصلی این پژوهش است، به تصمیم‌گیری ساده‌تر و آگاهانه‌تر خریداران سایت‌های تجارت الکترونیک منجر می‌شود. با توجه به اینکه تحلیل احساسات می‌تواند عقیده‌ها و نظرهای واقعی کاربران را دربارهٔ یک محصول استخراج کند، روند تصمیم‌گیری را بهبود می‌دهد. با پیشرفت چشمگیر کاربردهای متن‌کاوی در هوش مصنوعی و به‌خصوص پیشرفت شایان توجه شبکه‌های عمیق در این حوزه، استفاده از روش‌های نوین مبتنی بر شبکه‌های عمیق در تحلیل احساسات، به موضوع پژوهشی مهم برای پژوهشگران تبدیل شده است. از سوی دیگر، به‌دلیل خلأ موجود در این حوزه و اهمیت بسیار تحلیل احساسات برای کسب‌وکارهای تجارت الکترونیک، مسئلهٔ اصلی این پژوهش، ارائهٔ مدلی با تکیه بر شبکه‌های عمیق برای تحلیل احساسات خریداران، به‌منظور اصلاح فرایند تصمیم‌گیری و خرید است.

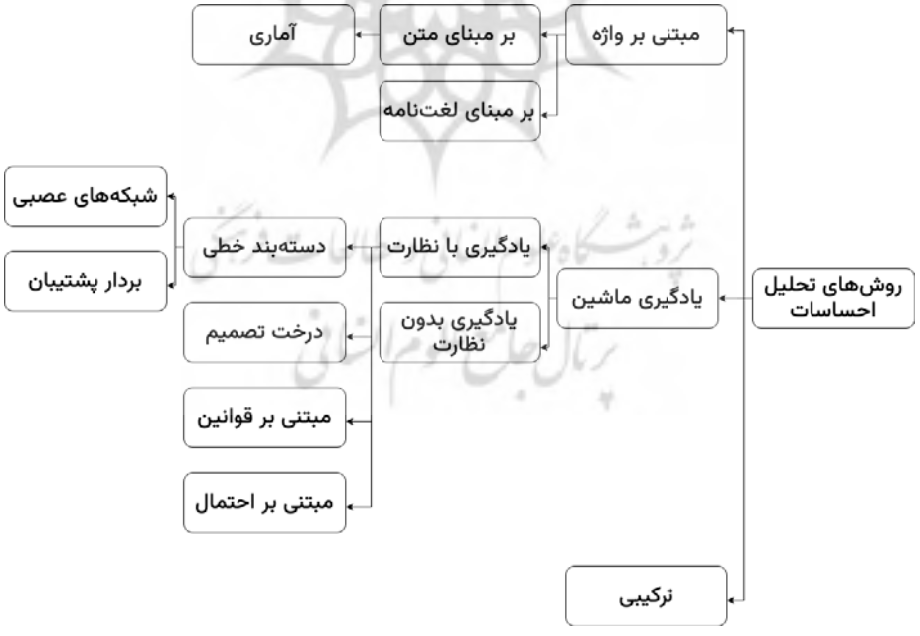
---

1. Internet World Stats
2. Murugavalli, Bagirathan, Saiprassanth & Arvindkumar
3. Liu
4. **Error! Hyperlink reference not valid.**
5. Convolutional Neural Network

پیشینهٔ نظری پژوهش

با طبقه‌بندی نظر مشتریان، می‌توان در سطح گسترده‌تری به قوت‌ها و ضعف‌های خدمات و محصولات از دیدگاه مشتریان پی برد. از طبقه‌بندی نظرهای مشتریان، برای شناخت سلیقهٔ مشتریان و هدفمندکردن تبلیغات استفاده می‌شود. همچنین می‌توان برای بخش‌بندی بازار و جهت‌دارکردن برنامه‌ریزی فروش، از طبقه‌بندی نظرهای مشتریان بهره برد.

امروزه حجم داده‌های موجود برای پردازش زیاد است و این حجم داده سالانه افزایش پیدا می‌کند (نیک‌نام و نیک‌نفس، ۱۳۹۵). تحلیل احساسات یکی از راه‌های استخراج اطلاعات از داده‌های متنی است (فنگ و ژان[1]، ۲۰۱۵). داده‌های متنی یکی از اشکال مهم داده‌های بدون ساختار[2] است. تحلیل احساسات به فرایند متن‌کاوی جهت استخراج لحن نوشته شدهٔ کاربر اشاره می‌کند که این لحن ممکن است مثبت باشد یا منفی (موروگاوالی و همکاران، ۲۰۱۷). به عبارتی دیگر با تجزیه‌وتحلیل عقیده‌ها، احساس‌ها، ارزیابی‌ها و نگرش مردم در خصوص موجودیت‌هایی همچون محصولات، خدمات، سازمان‌ها، افراد، مسائل، رویدادها، موضوعات و ویژگی‌های آن‌ها، احساسات افراد به هر یک مشخص می‌شود (لیو، ۲۰۱۵). تحلیل ادراکات از زبان طبیعی، چالش‌برانگیز است؛ زیرا به درک عمیقی از قوانین کار و ضمنی، منظم و نامنظم و نحوی و معنایی زبان نیاز دارد. تحقیقات تحلیل ادراکات با مشکلات حل‌نشدهٔ پردازش زبان طبیعی[3]، یعنی وضوح مسائلی مانند هم‌ارجاعی، تکرار و حساس به کلمه دست‌وپنجه نرم می‌کند (عباسی و خدیور، ۱۳۹۹). تکنیک‌های متنوعی جهت تحلیل احساسات وجود دارد که در شکل ۱ خلاصه‌ای از این روش‌ها مشاهده می‌شود.



شکل ۱. تکنیک‌های تحلیل احساسات

(پندی و همکاران[4]، ۲۰۱۷)

1. Fang and Zhan
2. Unstructured Data
3. Natural Language Processing
4. Panday, Rajpoot & Saraswat

رویکرد یادگیری ماشین[1]

این رویکرد به دو دستهٔ کلی تقسیم می‌شود:

۱. رویکرد یادگیری ماشین نظارت شده[2]: در رویکرد یادگیری با ماشین نظارت شده، خروجی داده‌های موجود به‌کمک یک ناظر بیرونی مشخص شده‌اند و سیستم تلاش می‌کند تا تابعی را بیابد که بتواند با مشاهدهٔ داده‌ها، خروجی مربوطه را تشخیص دهد. برای مثال با دادن فهرست نظرهای منفی و مثبت به سیستم، سیستم تلاش می‌کند تابعی را پیدا کند تا با مشاهدهٔ متن بتواند مثبت و منفی بودن لحن آن را تشخیص دهد (ونگ[3]، ۲۰۱۷)

۲. رویکردهای یادگیری بدون نظارت[4]: در این رویکرد، داده‌های مسئله خروجی مشخصی ندارند و سیستم برای یافتن الگویی میان داده‌ها تلاش می‌کند. برای مثال می‌توانیم با استفاده از رویکردهای یادگیری بدون ناظر، نظرهای یک کسب‌وکار را به دسته‌های کلی‌ای که از وجود آن‌ها در گذشته اطلاعی نداشتیم، تقسیم‌بندی کنیم و سپس متناسب با دسته‌های که بیشترین تعداد عضو را دارند، به اخذ تصمیم مناسب اقدام کنیم (لیو، ۲۰۱۵).

رویکرد واژه‌نامه[5]

در این رویکرد که رویکردی اساسی برای تجزیه‌وتحلیل متن است، به‌کمک لغت‌نامه‌ای که در آن، هر لغت با لحنی در ارتباط است و محاسبهٔ میانگین وزنی نمره‌های هر لحن، احساسات مربوط به متن مشخص می‌شود (ونگ، ۲۰۱۷).

رویکرد ترکیبی[6]

در بعضی از پژوهش‌ها، از روش ترکیبی استفاده شده است. در این رویکرد، داده‌هایی که بار احساسی آن‌ها با تکیه بر روش مبتنی بر لغت‌نامه تعیین می‌شوند، به‌عنوان داده‌های ورودی برای آموزش مدل در الگوریتم‌های یادگیری ماشین، در روش تحت نظارت استفاده می‌شوند (فیلهو و پاردو[7]، ۲۰۱۳).

رویکرد مبتنی بر یادگیری عمیق[8]

یکی از رویکردهای زیرمجموعهٔ یادگیری نظارت شده، رویکرد مبتنی بر یادگیری عمیق است. یادگیری عمیق به‌عنوان یادگیری چندسطحی از بازنمایی[9] مختلف، با استفاده از ساختار سلسله‌مراتبی و از روی ویژگی‌های سطح پایین تعریف می‌شود. مهم‌ترین ویژگی این رویکرد، یادگیری بازنمایی به‌عنوان رویکرد استخراج خودکار ویژگی‌ها از ورودی‌های سطح پایین است (لی و دونگ[10]، ۲۰۱۴).

1. Machine Learning
2. Supervised Learning
3. Wang
4. Unsupervised Learning
5. Lexicon-based Approach
6. Hybrid Approach
7. Filho & Pardo
8. Deep Learning
9. Representation
10. Li Dong, Dong Yu

رویکرد استفاده شده در این پژوهش، بر یادگیری عمیق مبتنی است و با استفاده از شبکه‌های کانوولوشن، شبکه‌های عصبی حافظهٔ طولانی ـ کوتاه مدت[1] انجام شده است.

شبکه‌های عصبی کانوولوشن

شبکه‌های عصبی کانوولوشن (CNN) یکی از مهم‌ترین روش‌های یادگیری عمیق است که لایه‌های متعدد در آن، به‌شیوه‌ای جدید و مستحکم آموزش می‌بینند. این شبکه‌ها زیرمجموعهٔ شبکه‌های عصبی هستند و داده‌های ورودی را به لایه‌های سلسله‌وار می‌دهند و در هر لایه با اعمال فیلترهای متعدد، ویژگی‌های چشمگیری داده را استخراج می‌کنند (قیوم، أنور، اویس و مجید[2]، ۲۰۱۷). به‌طور کلی شبکه‌های عصبی از سه لایهٔ کانوولوشن، ادغام[3] و تمام متصل[4] تشکیل می‌شود (سزاوار، فرسی و محمدزاده، ۱۳۹۷).

شبکه‌های عصبی حافظهٔ طولانی کوتاه‌مدت

یکی از مشکلات شبکه‌های عصبی بازگشتی، ناتوانی آن‌ها در یادگیری وابستگی بلندمدت است. جهت برطرف‌کردن این مشکل، یک معماری برای شبکه‌های عصبی بازگشتی پیشنهاد شد که به معماری حافظه طولانی کوتاه‌مدت معروف است و نقصان فراموشی ذکر شده را جبران می‌کند (هاکریتر و اشمیت‌هابر[5]، ۱۹۹۷). از زمان معرفی معماری LSTM تاکنون، بهبودهایی در راستای افزایش کارایی روی آن انجام شده است. یکی از انواع معماری LSTM معماری موسوم به LSTM دو طرفه[6] است که در این پژوهش از آن استفاده شده است. ساختار یک بلوک LSTM سه دروازه ورودی، خروجی و فراموشی را شامل می‌شود. خروجی بلوک LSTM به‌صورت بازگشتی، به ورودی بلوک و ورودی سه دروازهٔ نام‌برده شده متصل می‌شود. توابع فعال‌سازی در دروازه‌های ورودی و خروجی، معمولاً تابع تانژانت هذلولی است و تابع فعال‌سازی در دروازهٔ فراموشی تابع سیگموید[7] است (عدالت، عزمی و باقری نژاد، ۱۳۹۹).

پیشینهٔ تجربی پژوهش

در سال‌های اخیر پژوهش‌ها و مطالعات زیادی در زمینهٔ تحلیل احساسات و نظرهای شخصی کاربران انجام شده است. نظرهای آنلاین مشتریان بالقوه، اطلاعات ارزشمندی برای مشتریان را دربردارد؛ به همین دلیل، این نظرها به عامل مهمی برای کمک به تصمیم‌گیری کاربران تبدیل شده‌اند (المطیری و عبداله[8]، ۲۰۲۰).

نجف‌زاده، راحتی قوچانی و قائمی (۱۳۹۷) با ارائهٔ چارچوب نیمه‌نظارتی مبتنی بر لغت‌نامه جهت تحلیل نظرهای فارسی، به برچسب‌گذاری ادراکات بدون نیاز به خبرهٔ انسانی اقدام کردند.

1. Long short-term memory (LSTM)
2. Qayyum, Anwar, Awais & Majid
3. Pooling
4. Fully Connected
5. Hochreiter and Schmidhuber.
6. Bidirectional LSTM
7. Sigmoid
8. Almutairi & Abdullah

عباسی، سهرابی، مانیان و خدیور (۱۳۹۷)، مدلی جهت دسته‌بندی ادراکات خریداران کتاب از سایت آمازون ارائه کردند. آن‌ها با استفاده از روش وزن‌دهی، ادراکات مدنظرشان را با دقت ۸۰ درصد استخراج کردند.

پیکری، یعقوبی و طاهری (۱۳۹۴) فوت مرتضی پاشایی را در شبکهٔ اجتماعی توییتر با کاربرد تکنیک متن‌کاوی و تحلیل محتوا بررسی کردند. در این پژوهش، تمام توییت‌های انتشاریافته در پنج مقوله بررسی شده است.

مورگاوالی و همکاران (۲۰۱۷) با ارائهٔ مدلی به تجزیه‌وتحلیل ادراکات افراد در زمینهٔ تجارت الکترونیک پرداختند. در این پژوهش الگوریتمی جهت تشخیص قطبیت ادراکات بازخورد کاربران دربارهٔ محصولات ارائه شده است.

ژنگ، رن، ژو و فیث[1] (۲۰۱۹) از ترکیب چندین لغت‌نامه جهت تحلیل ادراکات نظر کاربران در خصوص برنامه‌های موبایلی استفاده کردند. همچنین با استفاده از روش‌های آماری، ویژگی‌های نظرهای کاربران و ارزش ادراکات آن‌ها را تجزیه‌وتحلیل نمودند. نتایج این پژوهش نشان می‌دهد که انحراف استاندارد ارزش ادراکات، تغییرات ادراکات میان مردم را منعکس می‌کند.

توکلی گارماسه و رافع (۱۳۹۵) به ارائه روشی برای تحلیل نظرهای ذیل خبر، در سایت‌های خبری با توجه به متن خبر پرداختند. آن‌ها تلاش کردند که با استفاده از ویژگی‌های دستوری متون، مانند اسم، فعل و همچنین تحلیل بار احساسی جمله‌ها، رابطهٔ نظر با خبر و دیدگاه نویسندهٔ نظر را با توجه به موضوع متن خبر کشف کنند.

حقیقی، آقازاده، خداداد حسینی و غریبی (۱۳۹۸) با بررسی نظر کاربران شبکهٔ اجتماعی اینستاگرام در ذیل مطالب منتشرشده در صفحات مربوط به شرکت‌ها و برندهای فعال در صنعت ماءالشعیر ایران، به تحلیل این نظرها و استخراج ابعاد هوشمندی رقابتی، شامل هوشمندی بازاریابی، هوشمندی رقبا، هوشمندی فناورانه، هوشمندی راهبردی و هوشمندی اجتماعی و سلامت اقدام کردند. آن‌ها در پژوهش خود با تکیه بر نیروی انسانی، به تحلیل نظرها پرداختند که مدل پیشنهادی پژوهش جاری از عهدهٔ این کار برخواهد آمد.

موسوی و امیری عقدایی (۱۳۹۹) با استفاده از روش تخصیص پنهان دریکله[2]، نظرهای خریداران تلویزیون از دیجی‌کالا را تحلیل و بررسی کردند. تحقیق آن‌ها نشان می‌دهد که ۳۰ ویژگی در محصولات روی رضایت یا نارضایتی مشتریان تأثیرگذار است.

پورسعید، شجاعی و نیک‌نفس (۱۴۰۰) با بررسی و تحلیل متن صفحات اینستاگرام سه مقصد گردشگری داخلی و چهار مقصد گردشگری خارجی در بازهٔ زمانی یک‌ساله، ۹۱۲ پُست اینستاگرامی در این صفحات را به‌کمک شبکه‌های عصبی و بُردار پشتیبان خطی طبقه‌بندی کردند.

عباسی و همکاران (۱۳۹۹) با استفاده از الگوریتم پرسپترون چندلایه[3]، مدلی برای طبقه‌بندی نظرهای خریداران تلفن همراه در وبسایت دیجی‌کالا ارائه کردند و توانستند با دقت ۸۹ درصد به طبقه‌بندی داده‌ها بپردازند.

1. Zang, Ren, Zhu & Faith
2. Latent Dirichlet Allocation
3. Multilayer Perception (MLP)

پس از بررسی مقاله‌های فوق، به این نتیجه رسیدیم که تاکنون از شبکه‌های عمیق و به‌خصوص شبکه‌های پیچشی برای تحلیل نظرهای مشتریان استفاده نشده است و خلأ آن در این زمینه احساس می‌شود؛ لذا در این پژوهش به انجام این کار می‌پردازیم.

روش‌شناسی پژوهش

در این پژوهش پس از گردآوری و پیش‌پردازش داده‌ها، از شبکه‌های عمیق، شامل شبکهٔ پیچشی برای مدل‌سازی استفاده شد. جامعهٔ آماری پژوهش، خریداران تلفن همراه از وبسایت دیجی‌کالاست.

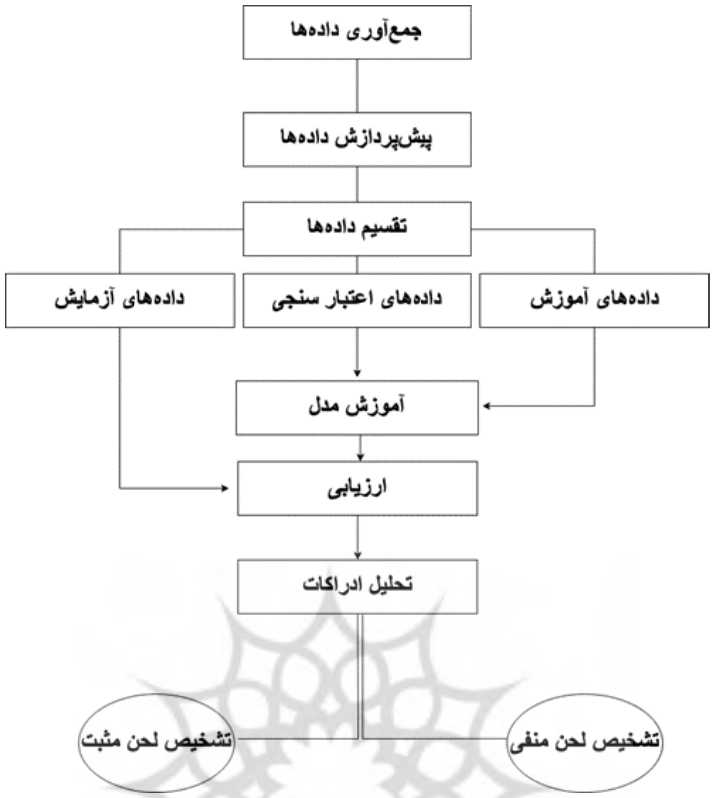
تعریف مسئله

امروزه هم‌راستا با افزایش کاربران، محتوایی که کاربران تولید می‌کنند هم افزایش می‌یابد. این محتوای تولید شده حاوی اطلاعات آشکار و پنهان است. با توجه به حجم و سرعت زیاد ثبت نظرها، نیاز به سیستمی هوشمند برای مدیریت و استخراج اطلاعات مهم از این نظرها محسوس است. در این پژوهش، مدل سریع و چابکی پیشنهاد داده‌ایم که ادراک کاربران را از نظرهایی که ثبت کرده‌اند، استخراج می‌کند. این مدل برای مدیران و کاربران وبسایت‌های تجارت الکترونیک بسیار کاربردی و حائز اهمیت است؛ زیرا هنگام ثبت نظر جدید در وبسایت‌های تجارت الکترونیک، مدیران این وبسایت‌ها از لحن نظر آگاه می‌شوند و به‌سرعت می‌توانند کاربران ناراضی را بیابند و در جهت رفع مشکلات آن‌ها تلاش کنند. به همین منظور، یک طبقه‌بند روی نظرهای خریداران تلفن همراه از وبسایت دیجی‌کالا با استفاده از رویکرد متن‌کاوی و شبکه‌های عمیق ارائه می‌شود.

مراحل حل مسئله

در مرحلهٔ نخست پژوهش، نظر خریداران تلفن همراه از دیجی‌کالا گردآوری شد. با توجه به اینکه متن این نظرها ممکن است حروف یا عبارات نامناسبی جهت پردازش در مراحل بعد باشد، بایستی برای استفاده در مراحل بعد، پیش‌پردازش شوند. مرحلهٔ پیش‌پردازش به‌طور کامل در ادامه تشریح می‌شود. لازم است که هر کلمه در متن نظرها به یک بُردار عددی تبدیل شود تا الگوریتم پیشنهادی که بردارهای عددی را طبقه‌بندی می‌کند، قابلیت استفاده در این مسئله را داشته باشد. برای تبدیل متن نظرها به بردارهای عددی، از مدل از پیش آموزش‌داده شدهٔ فست تکست[1] استفاده شده است. در مرحلهٔ بعد بردار کلمات را به سه مجموعه به نام‌های آموزش[2]، آزمایش[3] و اعتبارسنجی[4] تقسیم می‌کنیم. از مجموعه‌های آموزش و اعتبارسنجی در الگوریتم آموزشی استفاده خواهیم کرد، این الگوریتم در ادامه شرح داده شده است. سپس با استفاده از مجموعهٔ داده‌های آزمایش، عملکرد مدل پیشنهادی را ارزیابی و معیارهای صحت عملکرد را محاسبه می‌کنیم. در شکل ۲ روندها و رویکردهای هر کدام نشان داده شده است.

1. FastText.cc
2. Train set
3. Test set
4. Validation set



شکل ۲. نمودار روند حل مسئله

جمع‌آوری داده‌ها

جامعهٔ در دست بررسی، کاربران وبسایت دیجی‌کالاست که طی سال‌های ۱۳۹۴ تا ۱۳۹۵ از طریق این وبسایت به خرید تلفن همراه اقدام کرده و نظر خود را دربارهٔ کالای خریداری‌شده در وبسایت دیجی‌کالا ثبت کرده‌اند. این مجموعهٔ داده، از طریق وبسایت دیتاهارت[1] تهیه شده است. ویژگی داده‌ها در جدول ۱ نشان داده شده است.

جدول ۱. ویژگی‌های مجموعهٔ داده

ردیف	ویژگی	نوع ویژگی	توضیحات
۱	متن	متنی	متن نظر وارد شده توسط کاربر
۲	برچسب	دسته‌ای	برچسبی که مشخص می‌کند نظر کاربر مربوط به کدام ویژگی از تلفن همراه است
۳	لحن	دسته‌ای	برچسب مشخص‌کنندهٔ لحن نظر (مثبت یا منفی)

پیش‌پردازش داده‌ها

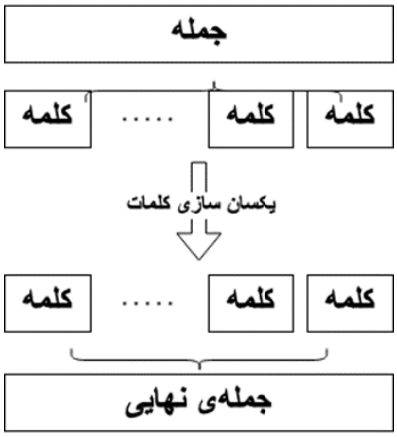
مجموعهٔ دادهٔ مورد بررسی در این پژوهش، ۱۱۵۴ نظر منحصربه‌فرد است که از این تعداد ۵۵۵ نظر با برچسب مثبت و ۵۹۹ نظر با برچسب منفی مشخص شدند. با توجه به اینکه تعداد نظرهای منفی بیشتر از نظرهای مثبت است، این

1. Dataheart.ir

احتمال وجود دارد که مدل پیشنهادی به‌سمت تعیین لحن منفی برای نظرهای مثبت گرایش پیدا کند. به بیان دیگر، ممکن است نظر مثبتی را به‌دلیل داشتن گرایش بیشتر به‌سمت نظرهای منفی، منفی تشخیص دهد. جهت افزایش دقت پیش‌بینی مدل پیشنهاد شده و جلوگیری از ایجاد تعصب به‌سمت طبقه‌های منفی، داده‌ها متعادل[1] شدند. به همین منظور تنها ۵۵۵ عدد از داده‌ها با برچسب منفی بررسی شدند. از سوی دیگر، برای اینکه متن نظرها همگی امکان پردازش داشته باشند، بایستی روی همهٔ آن‌ها اصلاحاتی اِعمال شود و تمامی متون به‌شکلی یکسان در الگوریتم آموزش استفاده شوند تا به بهترین شکل پردازش شوند. به همین منظور روی متن نظرها با استفاده از کتابخانهٔ هضم[2] در پایتون، اقدام‌های زیر انجام شد (خانهٔ بیگ‌دیتای ایران[3]).

۱. اصلاح انواع حرف «ک و ی» به معادل فارسی آنان.
۲. بررسی همزه و انواع مختلف املاهای موجود و اصلاح هر کدام (برای مثال تبدیل ؤ به و، ئ به ی، أ به ا، إ به ا و...)
۳. حذف شناسهٔ همزه از انتهای واژه‌هایی مثل شهداء و شناسه «آ» به «ا» مانند: آب به اب
۴. اصلاح نویسهٔ «طور» در واژه‌هایی مانند به طور، آن طور، این طور و...
۵. بررسی وجود حرف «ی» در انتهای لغاتی مانند خانهٔ ما و اصلاح آنان
۶. حذف تشدید از واژه‌ها
۷. تبدیل ارقام عربی و انگلیسی به معادل فارسی.
۸. اصلاح نویسهٔ نیم‌فاصله و اعراب و حذف فتحه، کسره و ضمه و همچنین تنوین‌ها
۹. حذف نویسهٔ «ـ» که برای کشش نویسه‌های چسبان استفاده می‌شود. مانند تبدیل«بـــــــر» و «بـــر» به «بر» و حذف نیم‌فاصله‌های تکراری
۱۰. چسباندن پسوندهای «تر»، «ترین» و... به آخر واژه‌ها
۱۱. اصلاح فاصله‌گذاری «ها» در انتهای واژه‌ها و همچنین پسوندهای «های»، «هایی»، «هایم»، «هایت»، «هایش» و اصلاح «می»، «نمی»، «درمی»، «برمی»، «بی» در ابتدای واژه‌ها
۱۲. تبدیل «ه» به «ه‌ی» و همچنین تبدیل «ب» متصل به ابتدای واژه‌ها به «به»
۱۳. اصلاح فاصله‌گذاری پسوندها
۱۴. حذف فاصله‌ها و نیم‌فاصله‌های اضافه به‌کار رفته در متن
۱۵. تصحیح فاصله‌گذاری برای علائم سجاوندی بدین صورت که علائم سجاوندی به لغات قبل از خود می‌چسبند و با لغت بعد از خود فاصله خواهند داشت.

1. Balance
2. Hazm
3. Bigdata-ir.com



شکل ۳ . پیش‌پردازش داده‌ها

همان گونه که در شکل ۳ ترسیم شده است، در ابتدا جمله‌ها به فهرستی از کلمات تبدیل شدند و تمامی ۱۵ مراحل یکسان‌سازی ذکر شده روی آن‌ها اجرا شد و کلمه‌ها با اتصال به یکدیگر، جمله‌های نهایی را تشکیل دادند و در نهایت این جمله‌ها به مرحلهٔ بَعد، یعنی مرحلهٔ تبدیل به بُردار، تحویل شدند.

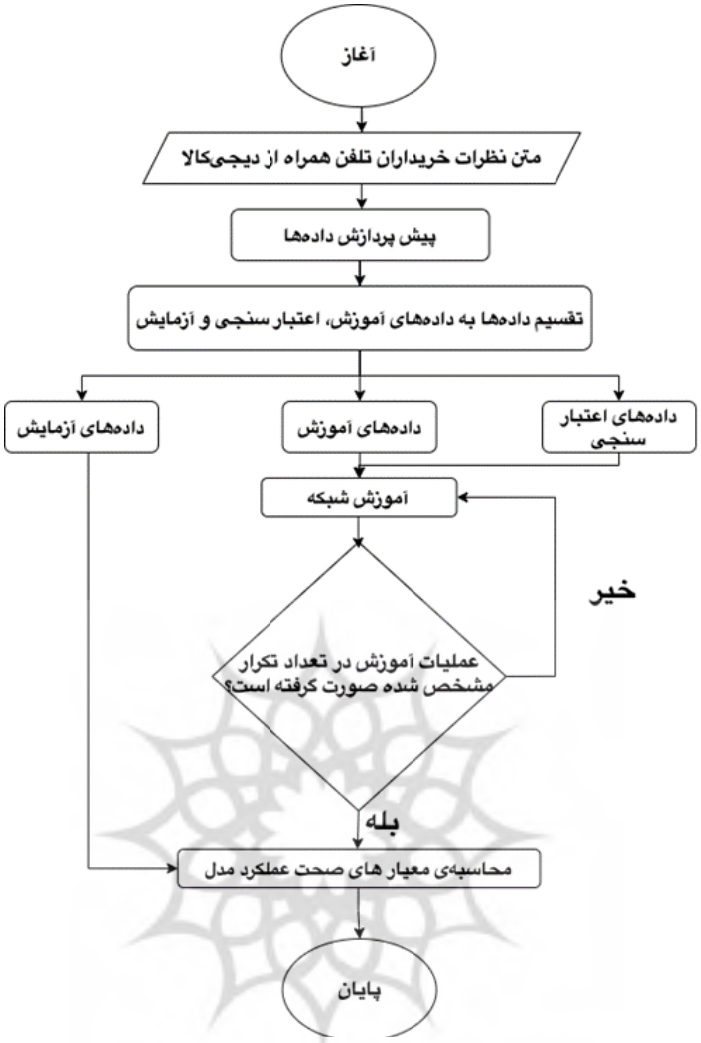
تبدیل متن به بردار

از آنجایی که مدل پیشنهادی پژوهش قادر است که بردارهای عددی را طبقه‌بندی کند، بایستی کلمه‌ها به اعدادی صحیح نگاشت شوند. برای این عمل، روش‌های مختلفی وجود دارد که به جهت بازیابی اطلاعات از داده‌های پیش‌پردازش‌شده و تبدیل متن به بردار، از روش تعبیه کلمه‌ها[1] استفاده شد. در این روش کلمات به شکل برداری از اعداد صحیح نمایش داده می‌شوند؛ به‌طوری که هم‌زمان اطلاعات نحوی و معنایی کلمه حفظ شود. روش تعبیهٔ کلمه‌ها، ابزاری بسیار قدرتمند است که در مسائل مدرن پردازش زبان طبیعی، به شکل وسیعی استفاده می‌شود (یو، ونگ، لای و ژنگ[2]، ۲۰۱۷). در این پژوهش برای تعبیه کلمه‌ها، از مدل از پیش‌آموزش‌دیدهٔ فست تکست استفاده شد. این مدل با کلمه‌های ویکی‌پدیای فارسی[3] آموزش دیده است و به‌خوبی کلمه‌های فارسی را به بردارهای متناظر آن‌ها نگاشت می‌کند.

الگوریتم

روندنمای[4] الگوریتم پیاده‌سازی و آموزش مدل معرفی‌شده در این پژوهش، در شکل ۳ به تصویر کشیده شده است. مدل پیشنهادی، ابتدا با استفاده از مجموعهٔ نظرهایی که برای آموزش و اعتبارسنجی در مرحلهٔ تقسیم داده‌ها جدا شدند، تعلیم داده می‌شود و پس از تعلیم مدل، می‌توان لحن نظرهای جدید را تشخیص داد.

1. Word Embedding
2. Yu, Wang, Lai & Xhang
3. Wikipedia
4. Flow chart



شکل ٤. روندنما روند حل مسئله

۱. داده‌های مسئله که متن نظرهای خریداران تلفن همراه بود، توسط تابع پیش‌پردازش داده‌ها، نرمال‌سازی و سپس به بردارهایی از اعداد حقیقی تبدیل شد.

۲. بردارهای نهایی همان گونه که در شکل ۵ نمایش داده شده است، به سه دسته به نام‌های آموزش، اعتبارسنجی و آزمایش تقسیم شدند.

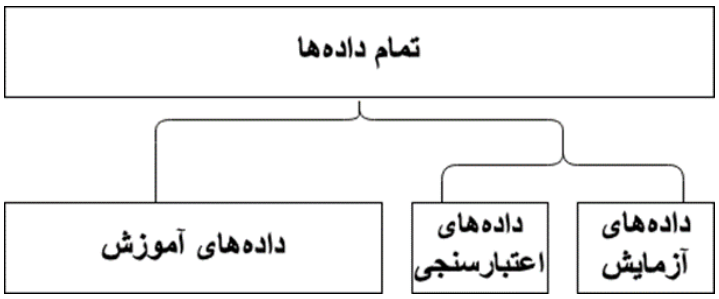
۳. مدل پیشنهادی پژوهش، یک شبکهٔ عمیق است و به‌کمک داده‌های آموزشی و اعتبارسنجی تعلیم داده می‌شود. به هر مرتبه اجرای الگوریتم آموزش روی شبکه، یک دورهٔ آموزش[1] گفته می‌شود. به تعداد دفعاتی مشخص، مدل تحت تعلیم دورهٔ آموزشی تشریح‌شده در زیر قرار گرفته است:

- در ابتدای دورهٔ آموزش، داده‌های آموزش به شکل دسته‌ای[2] به ورودی داده می‌شوند.
- خروجی لایه‌های شبکه، به‌ترتیب محاسبه‌شده و در نهایت خروجی شبکه محاسبه می‌شود.

1. Epoch
2. Batch

- به‌کمک الگوریتم گرادیان کاهشی تصادفی[1]، خطای شبکه محاسبه‌شده و از لایهٔ آخر تا لایهٔ ابتدایی جهت یادگیری و کاهش خطا در دورهٔ بعد آموزشی منتشر می‌شود و وزن‌های نرون‌های شبکه به‌روز می‌شوند.
- خطای مدل در هر دورهٔ آموزش توسط داده‌های اعتبارسنجی محاسبه می‌شود.

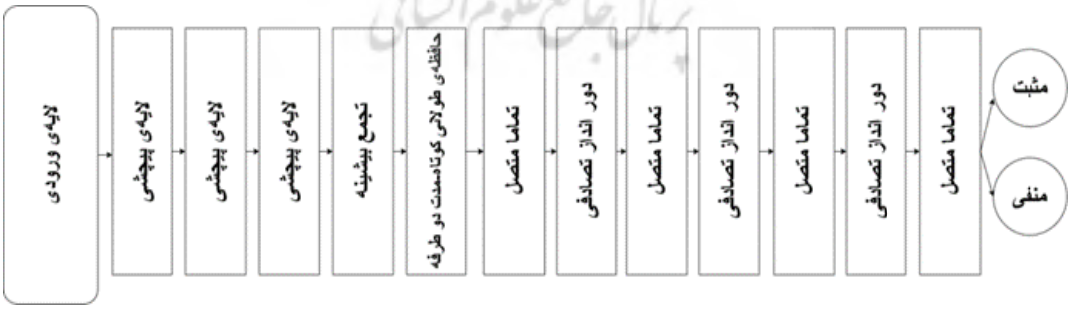
۴. پس از اتمام عملیات آموزش، معیارهای بررسی صحت عملکرد مدل توسط داده‌های آزمایش محاسبه می‌شوند.



شکل۵ . تقسیم داده‌ها به مجموعه‌های آزمایش، آزمون و آموزش

معماری مدل پیشنهادی

به‌منظور یافتن بهترین نتیجه، مدل‌های زیادی با معماری‌های متفاوت بررسی و آزمایش شد و از بین آن‌ها مدلی برگزیده شد که بهترین نتیجه را داشت. در پژوهش حاضر، از معماری مدل معرفی‌شدهٔ موجومدر، حسن، حسین و حسن[2] (۲۰۲۰) که برای طبقه‌بندی متن‌ها به زبان بنگلادشی استفاده کردند، به‌عنوان معماری پایه مدل استفاده شده است. پارامترهای مدل یادشده برای داده‌های این پژوهش و زبان فارسی بهینه‌سازی شدند. معماری مدل منتخب این پژوهش، از ۱۲ لایه تشکیل شده است عبارت‌اند از: لایهٔ پیچشی؛ یک لایهٔ تجمع بیشینه؛ یک لایهٔ حافظهٔ طولانی کوتاه‌مدت دوطرفه؛ سه‌جفت لایهٔ تمام متصل و لایهٔ دورانداز تصادفی؛ یک لایهٔ تمام متصل. برای بهینه‌سازی و آموزش شبکه، از روش بهینه‌سازی گرادیان کاهشی تصادفی با مقدار یادگیری[3] ۰/۰۰۱ استفاده شد. شکل ۶ معماری شبکه را نمایش می‌دهد.



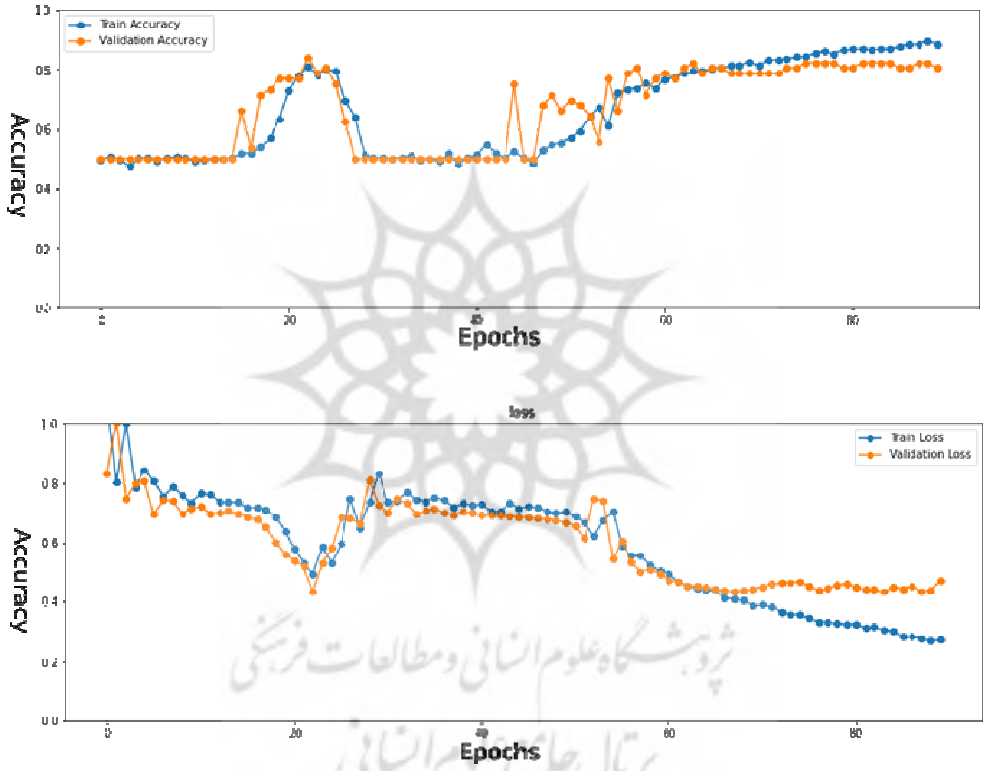
شکل٦. معماری شبکهٔ عمیق پیشنهاد شده

1. Stochastic Gradient Descent (SGD)
2. Mojumder, Hasan, Hossain & Hasan
3. Learning Rate

یافته‌های پژوهش

آموزش مدل

پس از مشخص‌شدن روش پژوهش و انجام پیش‌پردازش داده‌ها، داده‌های مسئله به داده‌هایی ساختاریافته برای استفاده در شبکه‌های عمیق جهت طبقه‌بندی داده‌ها تبدیل شد. ابتدا داده‌های مسئله به سه دسته داده‌های آموزش، داده‌های اعتبارسنجی و داده‌های آزمایش، به‌ترتیب با نسبت‌های ۸۰، ۱۰ و ۱۰ درصد تقسیم شدند. جهت پیاده‌سازی مدل پیشنهادی از کتابخانهٔ Tensorflow[1] استفاده شد. مدل پیشنهادی در این پژوهش ۹۰ دوره آموزش داده شد. شکل ۷، نمودار مقدار دقت و خطا[2] در هر دوره را روی داده‌های آموزش و اعتبارسنجی نمایش می‌دهد.



شکل ۷ . نمودار مقدار خطا و دقت در دوره‌های آموزش

ارزیابی

برای ارزیابی مدل‌های دسته‌بندی‌کننده، یک مجموعهٔ تصادفی از داده‌های مستقل از مجموعهٔ آموزش را به‌عنوان مجموعه آزمایش در نظر می‌گیریم. پس از آموزش مدل با مجموعهٔ داده‌های آموزش، مجموعهٔ آزمایش را دسته‌بندی می‌کنیم و برچسب‌های برآورد شده را با برچسب‌های واقعی مقایسه کرده و عملکرد مدل دسته‌بندی‌کننده را ارزیابی می‌کنیم (صنیعی آباده و محمودی، ۱۳۹۴). به‌منظور ارزیابی عملکرد مدل، پس از پیاده‌سازی آن، بایستی به‌کمک

1. Tensorflow.org
2. Loss

معیارهای بازخوانی[1] و دقت[2] و ارزیابی F، عملکرد آن را بررسی کرد. با استفاده از معیارهای پُرکاربرد دقت و بـازخوانی در حوزهٔ بازیابی اطلاعات، امکان محاسبهٔ میزان تناسب اسناد بازیابی شده توسط سیستم با نیاز کاربر فراهم می‌شود (راقاوان و جی وانگ[3]، ۱۹۸۹).

از ماتریس درهم‌ریختگی[4] برای محاسبهٔ مقادیر ارزیابی استفاده شده است. ماتریس درهـم‌ریختگـی عملکـرد مـدل مربوطه را نشان می‌دهد. در جدول ۲ ماتریس درهم‌ریختگی نمایش داده شده است که بر اساس اطلاعات ایـن مـاتریس ارزیابی صورت می‌گیرد.

جدول ۲. مقادیر ماتریس درهم‌ریختگی

		مقدار پیش‌بینی شده	
		مثبت	منفی
مقدار واقعی	مثبت	مثبت درست	مثبت نادرست
	منفی	منفی نادرست	منفی درست

به داده‌هایی که به‌شکل صحیح، به‌عنوان دستهٔ مثبت تشخیص داده شده‌اند، مثبت درست (TP) و به آن‌هایی که به اشتباه به‌عنوان دستهٔ نادرست تشخیص داده شده‌اند، مثبت نادرست (FP) می‌گوییم. همچنین به داده‌هایی که بـه شـکل درست به‌عنوان منفی مشخص شده‌اند، منفی درست (TN) و آن‌هایی که به اشتباه به‌عنوان منفی مشخص شده‌اند، منفی نادرست (EP) گفته می‌شود.

مقادیر ماتریس درهم‌ریختگی که برای مدل پیشنهادی این مقاله محاسبه شده است، در جدول ۳ مشاهده می‌شـود. به‌منظور محاسبهٔ مقادیر ارزیابی دقت، بازخوانی و معیار ارزیابی F از روابط زیر استفاده شده است.

رابطهٔ ۱) معیار دقت $\frac{TN}{TN + FN}$

رابطهٔ ۲) معیار بازخوانی $\frac{TN}{TN + FP}$

رابطهٔ ۳) معیار ارزیابی F $\frac{2 * Recall * Precision}{Recall + Precision}$

رابطهٔ ۴) معیار دقت $\frac{TN}{TN + FN}$

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Recall
2. Precision
3. Raghavan, Gwang
4. Confusion Matrix

رابطهٔ ۵) معیار میزان دقت $\frac{TN + TP}{TN + TP + FN + FP}$

رابطهٔ ۶) معیار میزان خطا $\frac{FN + FP}{TN + TP + FN + FP}$

جدول ۳. مقادیر ماتریس درهم‌ریختگی محاسبه شده برای مدل پیشنهادی

		مقدار پیش‌بینی شده	
		مثبت	منفی
مقدار واقعی	مثبت	۲۸	۰
	منفی	۴	۲۳

از طریق معیار بازخوانی، می‌توان دقت دسته‌بند را نسبت به کل داده‌ها ارزیابی کرد. معیار دقت، دقت دسته‌بند را با توجه به کل مواردی نشان می‌دهد که دسته‌بند پیش‌بینی کرده است. معیار بازخوانی کارایی دسته‌بند را نشان می‌دهد و معیار دقت، به‌طور کلی بر دقت پیش‌بینی دسته‌بندی‌کننده مبتنی است و نشان می‌دهد که به چه میزان می‌توان به خروجی دسته‌بندی‌کننده اعتماد کرد. معیار F که ترکیب دو معیار بازخوانی و دقت است، هنگامی استفاده می‌شود که نتوان اهمیت ویژه‌ای را برای هر یک از دو معیار بازخوانی و دقت نسبت به یکدیگر قائل شد. معیار میزان دقت که دقت دسته‌بندی‌کننده را ارزیابی می‌کند، نشان می‌دهد که دسته‌بند طراحی‌شده چند درصد از کل مجموعه رکوردهای آزمایشی را به‌درستی طبقه‌بندی می‌کند. معیار میزان خطا، برعکس میزان دقت است؛ بنابراین کمترین مقدار آن برابر صفر (بهترین کارایی) و بیشترین مقدار آن یک (ضعیف‌ترین کارایی) است (صنیعی آباده و محمودی، ۱۳۹۴). در این پژوهش چندین مدل آزمایش شد و از بین آن‌ها مدلی انتخاب شد که بهترین نتیجه را داشت. در اینجا فقط به ذکر نتیجهٔ آن می‌پردازیم. در جدول ۴ نتایج معیارهای ارزیابی محاسبه شده روی مدل پیشنهادی این پژوهش در مقایسه با رویکرد پیشنهادی عباسی و همکاران (۱۳۹۹) به نمایش درآمده است.

جدول ٤. مقادیر معیارهای ارزیابی محاسبه‌شده برای مسئله

طبقه‌بندی	دقت	بازخوانی	معیار F	مقدار دقت	مقدار خطا
مدل منتخب در این پژوهش	۰/۹۳	۱	۰/۹۳	۰/۸۸	۰/۱۲
مدل پیشنهادی عباسی و همکاران (۱۳۹۹)	۰/۸۹۲	۰/۸۹۲	۰/۸۹۲	۰/۸۹۲	۰/۱۰۸

همان گونه که مشاهده می‌شود، مدل این پژوهش در مقایسه با مدل پیشنهادی عباسی و همکاران (۱۳۹۹)، از لحاظ معیارهای دقت، بازخوانی و F عملکرد بهتری از خود نشان داده است.

طبقه‌بندی لحن نظرها

در این مرحله بعضی از نظرهای استخراج شده از وبسایت دیجی‌کالا با استفاده از مدل آموزش داده‌شدهٔ این پژوهش

طبقه‌بندی شده است. جدول ۵ نمونه‌ای از طبقهٔ پیشنهادی این نظرها را نشان می‌دهد. همان گونه که در جدول مشاهده می‌شود، مدل منتخب این پژوهش به‌خوبی می‌تواند لحن نظرها را استخراج کند.

جدول ۵. طبقه‌بندی بعضی از نظرهای خریداران

نظر	لحن تشخیص داده‌شده توسط مدل	لحن حقیقی
اپل گوشی خوش‌ساختی است.	مثبت	مثبت
بهبود کیفیت تصویر و هنگام برداشتن گوشی صفحه نمایش روشن می‌شود.	مثبت	مثبت
اصلاً گوشی خوبی نیست سیستم عامل تکراری دارد.	منفی	منفی
بیرون‌زدگی لنز دوربین در طراحی خوب نیست.	منفی	منفی
بهترین رابط کاربری در بین سایر برندهای اندرویدی را دارد.	مثبت	مثبت
همین 3D touch که خیلی‌ها می‌گویند که بی‌کاربرد است، برای من مفید بوده و به کارهای من راحتی بخشیده است.	مثبت	مثبت
اختلال در ارتباط با شبکه موبایل وقتی تنظیمات LTE فعال شده باشد.	مثبت	منفی
کسانی که این گوشی را از نظر سرعت از دیگر گوشی‌ها برتر می‌دانند، سخت در اشتباه هستند.	مثبت	منفی

بحث و نتیجه‌گیری

به‌طور کل در فضای بازاریابی دیجیتال و مجازی، برای تحلیل نظرهای مشتریان سعی می‌شود که از ابزارهای هوش مصنوعی استفاده شود و این تحلیل به سلیقه‌سازی، بازارسازی و بازارشناسی منجر می‌شود.

در پژوهش حاضر، یکی از کاربردهای مفید شبکه‌های پیچشی که از ابزارهای مدرن حوزهٔ هوش مصنوعی است، معرفی شد. این کاربرد از دیدگاه نظری و تئوریک، حائز اهمیت است. مدل پیشنهادی این پژوهش می‌تواند در تحلیل بازار، به‌خصوص بازار کسب‌وکارهای تجارت الکترونیک بسیار سودمند باشد. برای مثال با پیاده‌سازی در وبسایت‌های تجارت الکترونیک مانند دیجی‌کالا، به خریداران و مدیران کسب‌وکار کمک کند؛ برای مثال، می‌توان میزان کیفیت و رضایت مشتریان از محصولات و خدمات ارائه شده را با تحلیل تمامی نظرهای مربوط به آن‌ها سنجید. همچنین می‌توان با رصد شبکه‌های اجتماعی و پیاده‌سازی این مدل، هنگام ثبت دیدگاه منفی دربارهٔ محصول یا خدمات کسب‌وکار، به مدیران و کارکنان روابط عمومی آن کسب‌وکار جهت پیگیری و کسب رضایت مشتری ناراضی اطلاع‌رسانی کرد.

تاکنون روش‌های مختلف و نوآورانه‌ای جهت طبقه‌بندی نظرها و ایده‌ها معرفی شده است. یکی از بهترین روش‌های ارائه‌شده برای تحلیل نظرها و ادراک‌ها، روش عباسی و همکاران (۱۳۹۹) است که توانستند با دقت ۸۹ درصد لحن نظرها را استخراج کنند. در این پژوهش با رویکرد شبکه‌های عمیق و با استفاده از شبکه‌های پیچشی و حافظهٔ طولانی کوتاه‌مدت دوطرفه، پس از ۹۰ دورهٔ آموزش توانستیم با دقت ۹۳ درصد به طبقه‌بندی عقاید خریداران تلفن همراه در وبسایت دیجی‌کالا بپردازیم.

در این پژوهش از رویکردهای مدرن متن‌کاوی استفاده شد؛ پیشنهاد می‌شود که برای افزایش دقت و تعمیم‌پذیری مدل، در تحقیقات آینده به افزایش داده‌های مسئله یا استفاده از روش‌های تولید داده‌های مصنوعی اقدام شود. همچنین با افزایش داده‌های مسئله، می‌توان نظرهای ساختگی[1] را شناسایی و از ثبت آن‌ها جهت گمراهی کاربران جلوگیری کرد. به‌علاوه، نتایج تحلیل نظرها و ادراک‌های کاربران، منبع ارزشمندی است برای تصمیم‌گیری شرکت‌ها تا بتوانند استراتژی مناسبی به‌منظور ارائهٔ محصولات و خدمات جدید به مشتریان تدوین کنند. در تحقیقات آتی می‌توان با تفسیر و دریافت اطلاعات از صورتک‌های مفهومی[2]، به نتایج بهتری دست پیدا کرد؛ زیرا صورتک‌های مفهومی احساسات کاربران را به‌خوبی توصیف می‌کنند و می‌توانند میزان تنش، عصبانیت یا شادی کاربران را به شکل چشمگیری نشان دهند. استفاده از داده‌ها به زبان‌های دیگر، از جمله انگلیسی یا عربی می‌تواند دامنهٔ استفادهٔ تحقیق جاری را افزایش دهد. تحقیق روی سایر الگوریتم‌های هوش مصنوعی، می‌تواند زمینهٔ تحقیق جدیدی در ادامهٔ این پژوهش باشد تا امکان دقت و تصمیم‌گیری مدل پیشنهادی را افزایش دهد.

منابع

بخشی زاده، کبری؛ حاجی جعفر، علی و نصیری، حامد (۱۳۹۷). ترسیم نقشهٔ ذهنی مشتریان فروشگاه اینترنتی دیجی‌کالا با استفاده از تکنیک استخراج استعاره‌ای زالتمن (زیمت). *مدیریت بازرگانی*، ۱۰(۱)، ۴۹ – ۷۲.

پورسعید، محمدمهدی؛ شجاعی، فرزانه و نیک‌نفس، علی اکبر (۱۴۰۰). عوامل مؤثر بر برندسازی مکان با رویکرد داده‌کاوی (نمونه‌کاوی: شبکه اجتماعی اینستاگرام). *مدیریت بازرگانی*، ۱۳(۲)، ۴۷۳-۵۰۱.

پیکری، ناصر؛ یعقوبی، سید علی اصغر و طاهری، حامد (۱۳۹۴). تحلیل احساسات در شبکه اجتماعی توییتر با تکنیک متن کاوی. *کنفرانس بین‌المللی وب‌پژوهی*.

توکلی گارماسه، عاطفه و رافع، وحید (۱۳۹۵). ارائه روشی برای آنالیز احساسات در متن نظرات. *نخستین کنفرانس ملی تحقیقات بین رشته‌ای در مهندسی کامپیوتر، برق، مکانیک و مکاترونیک*.

حقیقی، محمد؛ آقازاده، هاشم؛ خداداد حسینی، سید حمید و غریبی، معین (۱۳۹۸). تبیین ابعاد هوشمندی رقابتی با بهره‌گیری از قابلیت‌های شبکه‌های اجتماعی در صنعت ماءالشعیر ایران. *مدیریت بازرگانی*، ۱۱(۴)، ۷۴۲-۷۶۱.

سزاوار، امیر؛ فرسی، حسن و محمدزاده، سجاد (۱۳۹۷). بازیابی تصویر مبتنی بر محتوا با استفاده از شبکه‌های عصبی کانولوشن عمیق. *مجله مهندسی برق دانشگاه تبریز*، ۴۸(۴)، ۱۵۹۵-۱۶۰۳.

صنیعی، محمد و محمودی، سینا (۱۳۹۴). *داده‌کاوی کاربردی*. تهران: نیازدانش.

عباسی، فاطمه؛ خدیور، آمنه و یزدی‌نژاد، محسن (۱۳۹۹). تحلیل ادراکات کاربران دربارهٔ خرید تلفن همراه در سایت دیجی‌کالا. *نشریه علمی مطالعات کسب‌وکار هوشمند*، ۸(۳۲)، ۱۸۱ – ۲۱۰.

1. Fake
2. Emoji

عباسی، فاطمه؛ سهرابی، بابک؛ مانیان، امیر و خدیور، آمنه (۱۳۹۷). ارائه مدلی جهت دسته‌بندی احساسات خریداران کتاب با استفاده از رویکرد ترکیبی. *مطالعات مدیریت کسب‌وکار هوشمند*، ۶(۲۱)، ۶۵-۹۲.

عدالت، محمد حسن؛ عزمی، رضا و باقری نژاد، جعفر (۱۳۹۹). بهبود دقت پیش‌بینی فرایندها در مدیریت فرایندهای کسب‌وکار با به‌کارگیری معماری LSTM. *چشم‌انداز مدیریت صنعتی*، ۱۰(۳)، ۷۱-۹۷.

موسوی، سید محسن؛ امیری عقدایی، سید فتح اله (۱۳۹۹). شناسایی عناصر سازنده «ارزش پیشنهادی به مشتری» و تأثیر آنها بر رضایت مشتری با استفاده از تحلیل احساسات بر مبنای متن‌کاوی. *مدیریت بازرگانی*، ۱۲(۴)، ۱۰۹۲-۱۱۱۶.

نجف‌زاده، محسن؛ راحتی قوچانی، سعید و قائمی، رضا (۱۳۹۷). یک چارچوب نیمه‌نظارتی مبتنی بر لغت‌نامه وفقی خودساخت جهت تحلیل نظرات فارسی. *پردازش علائم داده‌ها*، ۱۵(۲)، ۸۹ - ۱۰۲.

نیک‌نام، فرزاد و نیک‌نفس، علی‌اکبر (۱۳۹۵). بهبود روش‌های متن‌کاوی در کاربرد پیش‌بینی بازار با استفاده از الگوریتم‌های انتخاب نمونه اولیه. *مدیریت فناوری اطلاعات*، ۸(۲)، ۴۱۵ - ۴۳۴.

یکسان‌سازی یا نرمال سازی متون فارسی با استفاده از کتابخانه JHazm. (۱۴۰۰). دریافت شده در ۱۶ ۸، ۱۴۰۰، از https://bigdata-ir.com/?p=2387

References

Abbasi, F., Khadivar, A. & Yazdinezhad, M. (2020). Sentiment analysis of the Digikala Iphone buyers reviews. *IT Management Studies*, 8(32), 181-210. (*in Persian*)

Abbasi, F., Sohrabi, B., Khadivar, A. & Yazdinezhad, M. (2019). Presenting a model to classify book buyers using hybrid method. *IT Management Studies*, 8(32), 181 - 210. (*in Persian*)

Almutairi, Y., & Abdullah, M. (2020). IRHM: Inclusive Review Helpfulness Model for Review Helpfulness Prediction in E-commerce Platform. *Journal of Information Technology Management*, 12(2), 184-197.

Bakhshizadeh Borj, K., Haji Jafar, A., & Nasiri, H. (2018). Eliciting Mental Map of the Customers of Digikala E-Stores Using Zaltman Metaphor Elicitation Technique (ZMET). *Journal of Business Management*, *10*(1), 49-72. (*in Persian*)

Deng, L. & Yu, D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197–387. doi:10.1561/2000000039

Edalat, M. H., Azmi, R. & Bagherinezhad, J. (2020). An Enhanced LSTM Method to Improve the Accuracy of the Business Process Prediction. *Industrial Management Perspective*, 10(3), 71-97. (*in Persian*)

Fang, X. & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(5), 1 - 14.

Filho, P. & Pardo, T. (2013). An Improved Hybrid System for Sentiment Analysis in Twitter Messages. *Second Joint Conference on Lexical and Computational Semantics* (*SEM), 568-572.

Haghighi, M., Aghazadeh, H., Khodadad Hosseini, S., Gharibi, M. (2019). Explaining the Dimension of Competitive Intelligence through Utilizing Social Media Capabilities in Iran Non-Alcoholic Beverage Industry. *Journal of Business Management*, 11(4), 742-761. (*in Persian*)

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, 9(8), 1735–1780.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Synthesis Lectures on Human Language Technologies.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Mojumder, P., Hasan, M., Hossain, Md. F., & Hasan, K. M. A. (2020). A Study of fastText Word Embedding Effects in Document Classification in Bangla Language. *In T. Bhuiyan, Md. M. Rahman, & Md. A. Ali (Eds.), Cyber Security and Computer Science* (pp. 441–453). Springer International Publishing.

Mousavi, S., Amiri Aghdaie, S. (2021). Identifying the Constructive Elements of "Value Proposition" and their Impact on Customers' Satisfaction using Sentiment Analysis based on Text Mining. *Journal of Business Management*, 12(4), 1092-1116. doi: 10.22059/jibm.2020.302987.3847. (*in Persian*)

Murugavalli, S., Bagirathan, U., Saiprassanth, R., & Arvindkumar, S. (2017). Feedback analysis using Sentiment Analysis for E- commerce. *International Journal of Latest Engineering Research and Applications (IJLERA)*, 2(3), 84-90.

Najafzadeh, M., Rahati Ghouchani, S. & Ghaemi, R. (2018). A Semi-supervised Framework Based on Self-constructed Adaptive Lexicon for Persian Sentiment Analysis. *Signal and Data Processing*, 15(2), 89 - 102. (*in Persian*)

Niknam, F., & Niknafs, A. (2016). Improving Text Mining Methods in Market Prediction via Prototype Selection Algorithms. *Journal of Information Technology Management (JITM)*, 8(2), 415 - 434. (*in Persian*)

Pandey, A. C., Rajpoot, D. S. & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), 764 - 799.

Peykari, N., Yaghoubi, S. A. & Taheri, H. (2015). Sentiment analysis in twitter using data mining method. *International Conference on Web Research*. (*in Persian*)

Poursaeed, M., Shojaee, F., Niknafs, A. (2021). The Factors Affecting Place Branding based on Data Mining Approach (Case Study: Instagram Social Media). Journal of Business Management, 13(2), 473-501. doi: 10.22059/jibm.2021.309977.3944. (*in Persian*)

Qayyum, A., Anwar, S. M., Awais, M. & Majid, M. (2017). Medical Image Retrieval using Deep Convolutional Neural Network. *Neurocomputing*, 266, 8-20.

Raghavan, V., Gwang, J. (1989). A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance. *ACM Transactions on Information Systems*, 7(3), 205-229.

Saniee Abade, M., Mahmoudi, S. & Taherparvar, M. (2015). Applied data mining. (*in Persian*)

Sezavar, A., Farsi, H. & Mohammadzadeh, S. (2019). Content-Based Image Retrieval using Deep Convolutional Neural Networks. *Tabriz Journal of Electrical Engineering (TJEE)*, 48(4), 1595-1603. (*in Persian*)

Tavakoli garmase, A. & Rafe, V. (2016). Presenting a method to analyze sentiments in reveiws texts. *First National Conference on Industrial Engineering & Systems*. (*in Persian*)

Wang, Z. (2017). *The Evaluation of Ensemble Sentiment Classification Approach on Airline Services Using Twitter*. Masters dissertation, Technological University Dublin, 2017. doi:10.21427/D7190M.

Yu, L., Wang, J., Lai, K., Xhang, X,. (2017) Refining Word Embeddings for Sentiment Analysis. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Zhang, Y., Ren, W., Zhu, T. & Faith, E. (2019). MoSa: A Modeling and Sentiment Analysis System for Mobile Application Big Data. *Symmetry*, 11(1), 115.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی