دو فصلنامه علمی
فناوری اطلاعات و ارتباطات ایران

سال پانزدهم، شماره ۵۷ و۵۸ ، پاییز و زمستان ۱۴۰۲
صفحات:۱۲۲ الی ۱۳۷

Applying deep learning for improving the results of Persian comments' sentiment analysis in online retail stores

Faezeh Forootan*, Mohammad Rabiei**

*Ph.D. student in Information Technology Engineering, Faculty of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran

**Assistant Professor, Faculty of Electrical and Computer Engineering and Mechanics, Eivanekei Non-Profit University, Eivanekei, Iran

Abstract

The retail market industry is one of the industries that affects the economies of countries, the life of which depends on the level of satisfaction and trust of customers to buy from these markets. In such a situation, the retail market industry is trying to provide conditions for customer feedback and interaction with retailers based on web pages and online platforms. Because the analysis of published opinions play a role not only in determining customer satisfaction but also in improving products. Therefore, in recent years, sentiment analysis techniques in order to analyze and summarize opinions, has been considered by researchers in various fields, especially the retail market industry.

Therefore, in this study, with the aim of improving the results of extracting features from the text of Persian comments and increasing the accuracy of Persian sentiment analysis, a new framework for sentiment analysis at sentence level, based on BERT, CNN-BiLSTM feature extraction model and XGBoost classification model is proposed. Finally, the research results show 95.02% accuracy for classifying the sentiment of the text of Persian comments; Based on the proposed framework, according to which it can be acknowledged, CNN-BiLSTM is one of the powerful methods in extracting features from Persian text, which, while accurately extracting features, also increases increases the accuracy of Persian sentiment analysis.

Keywords: Sentiment Analysis, Persian Language, CNN-BiLSTM, BERT, Retail Market

به کارگیری یادگیری عمیق برای بهبود نتایج تحلیل احساسات نظرات فارسی فروشگاه‌های خرده‌فروشی آنلاین

فائزه فروتن*×، محمد ربیعی**

*دانشجوی دکتری مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز

**استادیار، دانشکده مهندسی برق و کامپیوتر و مکانیک، دانشگاه غیرانتفاعی ایوانکی

تاریخ دریافت: ۱۴۰۱/۰۴/۰۳ تاریخ پذیرش: ۱۴۰۱/۱۰/۱۴

نوع مقاله: پژوهشی

چکیده

صنعت بازار خرده‌فروشی از جمله صنایع اثرگذار بر اقتصاد کشورها است که حیات آن وابسته به میزان رضایت و اعتماد مشتریان برای خرید از این بازارها می‌باشد. در چنین شرایطی صنعت بازار خرده‌فروشی در تلاش است تا بر اساس صفحات وب و پلتفرم‌های آنلاین شرایطی را برای ثبت نظرات و تعامل مشتریان با خرده‌فروشان فراهم آورد. زیرا تحلیل نظرات منتشر شده نه تنها در تعیین میزان رضایت مشتریان بلکه در بهبود و ارتقا محصولات نقش دارند. از این رو در سال‌های اخیر تکنیک‌های تحلیل احساسات به منظور تحلیل و خلاصه‌سازی نظرات، مورد توجه پژوهشگران در حوزه‌های مختلف به‌ویژه صنعت بازار خرده‌فروشی قرار گرفته است.

در این پژوهش با هدف بهبود در نتایج استخراج ویژگی‌ها از متن نظرات فارسی و افزایش دقت تحلیل احساسات فارسی، یک چارچوب جدید برای تحلیل احساسات در سطح جمله، بر اساس BERT، مدل استخراج ویژگی CNN-BiLSTM و مدل طبقه‌بندی XGBoost پیشنهاد شده است. در نهایت نتایج پژوهش دقت ۹۵.۰۲٪ را برای طبقه‌بندی احساسات متن نظرات فارسی بر اساس چارچوب پیشنهادی نشان می‌دهد که بر اساس آن می‌توان اذعان داشت، CNN-BiLSTM از جمله روش‌های قدرتمند در استخراج ویژگی‌ها از متن فارسی است که ضمن استخراج دقیق ویژگی‌ها، باعث افزایش دقت تحلیل احساسات فارسی نیز می‌گردد.

واژگان کلیدی: تحلیل احساسات، زبان فارسی، CNN-BiLSTM، BERT، بازار خرده‌فروشی

۱. مقدمه

ظهور ویروس کرونا و لزوم به حداقل رساندن حضور مردم در اماکن عمومی و بازارها برای کاهش انتشار کرونا، موجب گردیده تا صنعت بازار خرده‌فروشی، در طی چند سال اخیر به شکل گسترده‌تری، دستخوش تغییر و تحول گردد. به گونه‌ای که تقریبا تمام فرآیندهای این بازار برای ارائه خدماتی مانند فروش آنلاین محصولات و ایجاد شرایطی برای ثبت نظرات مشتریان، در زمانی کوتاه و با سرعتی چشمگیر تغییر پیدا کرده‌اند. در چنین شرایطی بازارهای خرده‌فروشی آنلاین در کشورهای مختلف به ویژه در کشور ایران به سرعت گسترش یافتند و با استقبال زیادی از سوی کاربران مواجه شدند. همین مسئله موجب گردید تا فروشگاه‌های خرده‌فروشی آنلاین در ایران با

×نویسنده مسئول: فائزه فروتن Forootan_faezeh@yahoo.com

عرضه اقلام و محصولات متنوع، باعث افزایش تمایل عموم مردم به ویژه قشر جوان نسبت به خرید آنلاین شوند. دیجی‌کالا از جمله وب‌سایت‌های خرده‌فروشی آنلاین در ایران است؛ که به واسطه فروش بیش از ۴ میلیون محصول متنوع و بازدید ماهانه ۳۰ میلیون کاربر یکتا، نقش مهمی را در تجارت آنلاین و اقتصاد دیجیتال ایران، ایفا می‌کند و هم اکنون به عنوان بزرگترین و محبوب‌ترین وب‌سایت خرده‌فروشی آنلاین در ایران و خاورمیانه محسوب می‌شود [۱-۲]. دیجی‌کالا در کنار عرضه محصولات، این امکان را برای کاربران فراهم آورده تا در طی فرآیند خرید، نظرات خود را به صورت متن با دیگران به اشتراک بگذارند. به همین دلیل در حال حاضر دیجی‌کالا به عنوان یکی از غنی‌ترین منابع متنی زبان فارسی در زمینه متن‌کاوی [۱] و پردازش زبان طبیعی [۲] محسوب می‌گردد؛ زیرا توانسته است در سال ۲۰۲۱-۲۰۲۰ بیش از ۱۲ میلیون و ۵۶۰ هزار نظر متنی کاربران در زمینه محصولات را، منتشر نماید [۱]. تحلیل این داده‌های متنی و دست‌یابی به نتایج آن‌ها برای حوزه‌های مختلفی با ارزش و حیاتی است. به همین دلیل پژوهشگران در حوزه‌های مختلف اقتصادی، تجاری، بازاریابی و غیره در تلاشند تا از تکنیک‌های تحلیل احساسات [۳]، که به عنوان تحلیل ذهنیت [۴] نیز شناخته می‌شود، برای خلاصه‌سازی، تحلیل و طبقه‌بندی نظرات استفاده کنند [۲]. این در حالیست که تحلیل نظرات فارسی به علت کمبود ابزارهای پردازشی برای این زبان، به عنوان یکی از دشوارترین کارها در زمینه تحلیل احساسات، محسوب می‌گردد. از این رو پژوهشگران در حوزه زبان فارسی، همواره در تلاشند؛ تا با استفاده از روش‌ها و تکنیک‌های جدید تحلیل احساسات، به نتایج معتبری در این زمینه دست یابند.

در این پژوهش با هدف بهبود در نتایج استخراج ویژگی‌ها از متن نظرات فارسی و افزایش دقت تحلیل احساسات، چارچوبی در سطح جمله و بر اساس روش تعبیه‌سازی [۵] BERT، مدل استخراج ویژگی [۶] CNN-BiLSTM و الگوریتم XGBoost، پیشنهاد شده است. بر همین اساس مهمترین نوآوری‌های این چارچوب عبارتند از:

۱. به کارگیری مدل ترکیبی یادگیری عمیق CNN-BiLSTM با هدف بهبود نتایج در مرحله استخراج ویژگی‌ها از متن نظرات فارسی.
۲. به کارگیری مدل طبقه‌بندی XGBoost، در حالیکه استخراج ویژگی‌ها بر اساس CNN-BiLSTM انجام گردیده است.
۳. مقایسه عملکرد و نتایج مدل‌های CNN-BiLSTM و CNN-LSTM، در حالیکه برای استخراج ویژگی‌ها از متن نظرات فارسی استفاده شده‌اند.

لازم به ذکر است که تا به امروز، هیچ یک از پژوهشگران فارسی زبان، از موارد فوق برای تحلیل احساسات متون نظرات فارسی استفاده نکرده‌اند. همچنین ساختار مقاله در ادامه بدین صورت تنظیم شده است:

در بخش دوم با عنوان "کارهای مرتبط"، پژوهش‌های صورت گرفته در حوزه تحلیل احساسات بررسی شده‌اند. مراحل چارچوب پیشنهادی و نتایج هر یک از این مراحل، به ترتیب در بخش‌های سوم و چهارم با عنوان‌های "چارچوب پیشنهادی" و "پیاده‌سازی و تحلیل نتایج" مطرح شده‌اند. در نهایت مقاله براساس بخش پنجم و با عنوان "نتیجه‌گیری" به پایان رسیده است.

۲. کارهای مرتبط

تا به امروز، پژوهش‌های بسیاری در زمینه تحلیل احساسات و بر اساس روش‌های مختلف، صورت پذیرفته است. با این حال آنچه امروزه از سوی پژوهشگران، بیش از پیش مورد توجه قرار گرفته است؛ استفاده از مدل‌های ترکیبی یادگیری عمیق [۷] برای بهبود عملکرد تحلیل احساسات می‌باشد. از این رو با توجه به هدف اصلی پژوهش، مبنی بر استفاده از مدل CNN-BiLSTM برای بهبود عملکرد استخراج ویژگی‌ها، در این بخش تنها به بررسی کارهایی پرداخته شده است، که از مدل ترکیبی CNN-BiLSTM به منظور بهبود عملکرد در استخراج ویژگی‌ها و طبقه‌بندی احساسات استفاده کرده‌اند.

۲-۱. پژوهش‌های صورت گرفته بر اساس CNN-BiLSTM در مرحله طبقه‌بندی احساسات

شن و همکاران [۸] در ۲۰۱۷ برای بهبود دقت طبقه‌بندی احساسات از CNN-BiLSTM، استفاده کرده‌اند [۳]. آن‌ها بعد از پیش‌پردازش مجموعه نظرات فیلم IMDB، از روش GloVe برای تبدیل نظرات به بردارهای عددی استفاده کرده‌اند. سپس CNN-BiLSTM را بر روی نظرات اعمال کرده‌اند، که در نهایت نتایج نشان می‌دهد روش پیشنهادی

[1] Text Mining
[2] Natural Language Processing
[3] Sentiment Analysis
[4] Subjective Analysis
[5] Word Embedding
[6] Convolutional Neural Network-Bidirectional LSTM
[7] Hybrid deep learning
[8] Shen et al.

آن‌ها توانسته است متون نظرات را با دقت ۸۹/۷۰٪ در دو گروه مثبت و منفی طبقه‌بندی نماید.

رانویی و همکاران [1] در ۲۰۱۹ بعد از پیش‌پردازش نظرات پیرامون محصولات، از روش Doc2vec برای تعبیه‌سازی استفاده کرده‌اند [۴]. سپس مدل CNN-BiLSTM را برای طبقه‌بندی نظرات در سه گروه احساسات مثبت، منفی و خنثی به کار برده‌اند. آن‌ها در ادامه به مقایسه عملکرد روش پیشنهادی خود با مدل‌های CNN، LSTM، BiLSTM و CNN-LSTM پرداخته‌اند. نتایج نشان می‌دهد مدل CNN-BiLSTM با دقت ۹۰/۶۶٪ نسبت به دیگر مدل‌ها عملکرد بهتری را برای طبقه‌بندی احساسات داشته است.

سالور و ایدین [2] در ۲۰۲۰ برای تعبیه‌سازی توییت‌های پیرامون محصولات، از روش‌های FastText، جاسازی در سطح کاراکتر [3] و Word2Vec استفاده کرده‌اند [۵]. سپس هر یک از مدل‌های یادگیری عمیق CNN و LSTM، BiLSTM، GRU [4] و همچنین مدل ترکیبی CNN-BiLSTM را بر مبنای هر سه روش تعبیه‌سازی بر روی نظرات اعمال کرده‌اند؛ که در نهایت نتایج نشان می‌دهد CNN-BiLSTM بر مبنای FastText و با دقت ۸۲/۱۴٪، عملکرد بهتری را برای طبقه‌بندی نظرات در دو گروه مثبت و منفی داشته است.

لی و همکاران [5] در ۲۰۲۰ بعد از پیش‌پردازش توییت‌های گردشگران چینی، از روش Word2Vec برای تعبیه‌سازی استفاده کرده‌اند [۶]. سپس مدل‌های ترکیبی CNN-LSTM و CNN-BiLSTM را به صورت موازی بر روی نظرات اعمال کرده‌اند؛ که نتایج نشان می‌دهد CNN-BiLSTM با دقت ۹۵٪/۰۱ عملکرد مناسب‌تری را برای طبقه‌بندی احساسات داشته است.

پروتشا و همکاران [6] در ۲۰۲۲ به تحلیل نظراتی پرداختند که کاربران بنگالی زبان در شبکه‌های اجتماعی مختلف مانند توییتر، فیس‌بوک، وب‌سایت‌های فروش آنلاین منتشر کرده بودند [۷]. آن‌ها بعد از پیش‌پردازش نظرات از روش BERT برای تعبیه‌سازی و از مدل ترکیبی CNN-BiLSTM برای طبقه‌بندی و تحلیل احساسات در دو گروه مثبت و منفی استفاده کرده‌اند. آن‌ها در ادامه به مقایسه نتایج روش پیشنهادی خود با دیگر الگوریتم‌های یادگیری ماشین مانند ماشین بردار پشتیبان [7]، جنگل تصادفی [8]، رگرسیون خطی [9]، K-نزدیکترین همسایگی [10]، نیوبیز [11]، درخت تصمیم و همچنین الگوریتم‌های یادگیری عمیق CNN، ANN، LSTM بر اساس هر یک از روش‌های تعبیه‌سازی GloVe، FastText، Word2Vec پرداخته‌اند. در نهایت نتایج نشان می‌دهد روش پیشنهادی آن‌ها (مدل طبقه‌بندی CNN-BiLSTM بر اساس BERT) با دقت ۸۴/۹۳٪، نسبت به دیگر روش‌ها عملکرد مناسبی را بر روی نظرات زبان بنگالی داشته است.

آیتیران [12] در ۲۰۲۲ به تحلیل احساسات نظرات انگلیسی منتشر شده در زمینه رستوران‌ها و محصولات الکترونیکی سایت آمازون پرداخته است [۸]. او پس از اجرای مراحل پیش‌پردازش از GloVe برای تعبیه‌سازی، از CNN برای استخراج ویژگی‌ها و از مدل BiLSTM مبتنی بر مکانیزم توجه [13] برای طبقه‌بندی احساسات استفاده کرده است. نتایج نشان می‌دهد روش پیشنهادی او توانسته است متون نظرات را با دقت ۸۱/۹۶٪ در سه گروه مثبت، منفی و خنثی طبقه‌بندی نماید.

گیرادارهای و همکاران [14] در ۲۰۲۲ به تحلیل احساسات کلمات نفرت انگیزی که در توییتر منتشر شده بودند، پرداخته است [۹]. او بعد از پیش‌پردازش از GloVe برای تعبیه‌سازی و از مدل C-BiLSTM که ترکیبی از مدل‌های یادگیری عمیق CNN و BiLSTM می‌باشد، برای تحلیل احساسات استفاده کرده است. نتایج نشان می‌دهد روش او با دقت ۷۳/۵۱٪ توانسته است توییت‌ها را در ۴ گروه توهین‌آمیز [15]، عادی، نفرت، و هرزنامه طبقه‌بندی نماید.

دیپیکا و همکاران [16] در ۲۰۲۲ برای طبقه‌بندی احساسات موجود در متون شعرهای انگلیسی، پس از پیش‌پردازش از n-gram برای مرحله استخراج ویژگی‌ها استفاده کرده‌اند [۱۰]. سپس بر اساس مدل ترکیبی CNN-BiLSTM و مکانیزم توجه مراحل طبقه‌بندی و تحلیل احساسات را انجام داده‌اند. نتایج نشان می‌دهد روش پیشنهادی آن‌ها توانسته است متون شعر را با دقت ۹۲٪ در ۶ گروه خشم، ترس، شادی، عشق، غم، شگفتی [17] طبقه‌بندی نماید.

یو و همکاران [18] در ۲۰۲۲ به تحلیل نظرات چینی که در زمینه شهر هوشمند منتشر شده بودند، پرداخته‌اند [۱۱]. آن‌ها بعد از

[1] Rhanoui et al.
[2] Salur & Aydin.
[3] Character-level Embedding
[4] Gated Recurrent Units (GRU)
[5] Li et al.
[6] Prottasha
[7] Support Vector Machine (SVM)
[8] Random Forest Tree
[9] Linear Regression
[10] K-nearest neighbors
[11] Naive Bayes
[12] Ayetiran
[13] Attention Mechanism
[14] Giradharbhai et al.
[15] Abusive
[16] Deepika et al.
[17] Surprise
[18] Yue et al.

پیش‌پردازش از روش Word2Vec برای تبدیل نظرات به بردارهای عددی استفاده کرده‌اند. سپس الگوریتم تخصیص پنهان دیریکله [1] را برای طبقه‌بندی موضوعات مطرح شده در زمینه شهر هوشمند و مدل ترکیبی CNN-BiLSTM را برای طبقه‌بندی احساسات به کار گرفته‌اند. نتایج آن‌ها در زمینه طبقه‌بندی موضوعات نشان می‌دهد که، بحث‌های عمومی در مورد شهرهای هوشمند کلان محور بوده و بر طرح‌بندی استراتژیک و کاربردهای فنی تمرکز دارد. همچنین نتایج نشان می‌دهد روش پیشنهادی آن‌ها توانسته است با دقت ۹۴/۱۹٪ نظرات را در سه گروه مثبت و منفی و خنثی طبقه‌بندی نماید.

اما با توجه به پژوهش‌های اشاره شده در دیگر زبان‌ها، تا به امروز در هیچ یک از پژوهش‌های صورت گرفته بر روی متن نظرات فارسی، از CNN-BiLSTM برای طبقه‌بندی احساسات استفاده نشده است. با این حال در ادامه به آن دسته از پژوهش‌های فارسی اشاره گردیده که، از مدل‌های CNN، BiLSTM (به صورت جداگانه و یا ترکیبی) و یا از مدل ترکیبی CNN-LSTM برای طبقه‌بندی احساسات استفاده کرده‌اند و به نتایج معتبری در این زمینه دست یافته‌اند.

شرامی و همکاران [2] در ۲۰۲۰ به تحلیل نظرات فارسی پیرامون محصولات دیجی‌کالا پرداخته‌اند و از FastText و Keras Embedding برای تبدیل نظرات به بردارهای عددی استفاده کرده‌اند [۱۲]. آن‌ها در ادامه، هر یک از مدل‌های CNN و Bi-LSTM را به صورت جداگانه، برای طبقه‌بندی دودویی [3] و طبقه‌بندی چندکلاسه [4]، بر روی نظرات اعمال کرده‌اند که نتایج نشان می‌دهد، Bi-LSTM با ۶۹/۳۳٪ = F1_score در طبقه‌بندی دودویی و ۹۱/۸۹٪ = F1_score در طبقه‌بندی چند کلاسه؛ عملکرد بهتری را نسب به CNN داشته است.

شمالی و همکاران [5] در ۲۰۲۱، به تحلیل نظرات فارسی پیرامون محصولات دیجی‌کالا پرداخته‌اند و از TF-IDF [6] و FastText برای تبدیل نظرات به بردارهای عددی استفاده کرده‌اند [۱۳]. سپس مدل‌های یادگیری عمیق CNN و Bi-LSTM را به صورت جداگانه و بر اساس FastText و همچنین مدل‌های یادگیری ماشین نیوبیز و رگرسیون خطی را بر اساس TF-IDF بر روی نظرات پیاده‌سازی کرده‌اند. در نهایت نتایج آن‌ها نشان می‌دهد CNN بر اساس FastText و با دقت ۹۹/۶٪ نسبت به دیگر الگوریتم‌ها، عملکرد مناسبی را برای طبقه‌بندی احساسات در دو گروه مثبت و منفی داشته است.

دشتی‌پور و همکاران [7] در ۲۰۲۱ برای طبقه‌بندی احساسات فارسی، هر یک از مدل‌های CNN و Stacked-BiLSTM را بر اساس روش FastText بر روی نظرات پیرامون فیلم اعمال کرده‌اند [۱۴]. نتایج نهایی آن‌ها نشان می‌دهد Stacked-BiLSTM با دقت ۹۶٪ عملکرد بهتری را نسبت به CNN داشته است.

منصور و همکاران در ۱۴۰۰ با هدف تحلیل احساسات نظرات بورس فارسی از مدل ترکیبی CNN-LSTM برای طبقه‌بندی چند کلاسه استفاده کرده‌اند [۱۵]. نتایج آن‌ها نشان می‌دهد مدل CNN-LSTM با دقت ۷۸/۹۰٪، عملکرد مناسبی را برای تحلیل احساسات داده‌های بورس داشته است.

۲-۲. پژوهش‌های صورت گرفته بر اساس CNN-BiLSTM برای استخراج ویژگی‌ها

یون و کیم [8] در ۲۰۱۷، برای تحلیل احساسات توییت‌های پیرامون محصولات، از Word2Vec برای تعبیه‌سازی نظرات استفاده کرده‌اند [۱۶]. سپس مدل استخراج ویژگی CNN-BiLSTM را بر روی نظرات پیاده‌سازی کرده‌اند و در نهایت طبقه‌بندی احساسات را بر اساس لایه تماما متصل [9]، در سه گروه مثبت و منفی و خنثی انجام داده‌اند. نتایج آن‌ها نشان می‌دهد روش پیشنهادی با دقت ۷۰/۰۸٪ دارای عملکرد مناسبی برای تحلیل احساسات بوده است.

منگ و همکاران [10] در ۲۰۱۹ به تحلیل نظرات پیرامون محصولات و رستوران‌ها پرداخته‌اند [۱۷]. آن‌ها برای استخراج ویژگی‌ها از CNN-BiLSTM استفاده کرده‌اند و سپس مکانیزم توجه را برای در نظر گرفتن ارتباط میان ویژگی‌ها و طبقه‌بندی احساسات در دو گروه مثبت و منفی به کار برده‌اند. نتایج آن‌ها نشان می‌دهد استفاده از CNN-BiLSTM باعث شده است تا عملکرد مدل طبقه‌بندی بهبود یابد و به دقت ۸۳/۲۱ ٪ دست یابد.

امبابی و همکاران [11] در ۲۰۲۰ برای تحلیل نظرات عربی پیرامون هتل از FastText برای تعبیه‌سازی استفاده کرده‌اند [5]. سپس با هدف بهبود عملکرد در مرحله استخراج ویژگی‌ها، CNN-LSTM را به کار گرفته‌اند و در ادامه الگوریتم SVM [12] را برای طبقه‌بندی احساسات استفاده

[1] Linear Discriminant Analysis
[2] Sharami et al.
[3] Binary-classification
[4] Multi-classification
[5] Shumaly et al.
[6] Term Frequency-Inverse Document Frequency (TF-IDF)
[7] Dashtipour et al
[8] Yoon and Kim.
[9] Fully Connected
[10] Meng et al.
[11] Ombabi et al.
[12] Support Vector Machines (SVM)

کرده‌اند. نتایج نهایی آن‌ها نشان می‌دهد استفاده از CNN-LSTM باعث بهبود عملکرد طبقه‌بندی و تحلیل احساسات با دقت ۹۰/۷۵٪ شده است.

گن و همکاران[1] در ۲۰۲۱ برای بهبود در استخراج ویژگی‌ها از CNN-BiLSTM استفاده کرده‌اند [۱۸]. سپس طبقه‌بندی احساسات در دو گروه مثبت و منفی را بر اساس مکانیزم توجه انجام داده‌اند. نتایج نهایی آن‌ها نشان می‌دهد استفاده از CNN-BiLSTM برای استخراج ویژگی‌ها باعث افزایش و بهبود عملکرد طبقه‌بندی احساسات و دست‌یابی به دقت ۹۸٪/۲۸ بر روی نظرات پیرامون هتل شده است.

اما در زمینه کارهای فارسی، تنها می‌توان به یکی دیگر از پژوهش‌های دشتی‌پور و همکاران[2] در سال ۲۰۲۱ اشاره نمود که بعد از تعبیه‌سازی نظرات بر اساس FastText، مدل Bi-LSTM را برای استخراج ویژگی‌ها به کار برده‌اند [۱۹]. در ادامه برای تحلیل احساسات چندوجهی[3] مبتنی بر متن و تحلیل احساسات تک وجهی[4] نظرات پیرامون محصولات دیجی‌کالا، از مدل طبقه‌بندی LSTM + Dependency-based استفاده کرده‌اند. در نهایت نتایج، دقت ۸۸/۰۱٪ را برای روش آن‌ها در زمینه تحلیل احساسات تک وجهی و دقت ۹۱٪/۳۹ را برای تحلیل احساسات چند وجهی بر اساس Late Fusion (decision-level) نشان می‌دهد.

بنابراین با توجه به نتایج پژوهش‌های اشاره شده که حاکی از بهبود عملکرد استخراج ویژگی‌ها و طبقه‌بندی احساسات، بر اساس مدل CNN-BiLSTM می‌باشد؛ در این پژوهش از این مدل ترکیبی با هدف بهبود عملکرد استخراج ویژگی‌ها از متن نظرات فارسی استفاده شده است. این در حالیست که تا به امروز در هیچ یک از پژوهش‌های فارسی از این روش برای استخراج ویژگی‌ها استفاده نشده است.

۳. چارچوب پیشنهادی

هدف اصلی در این پژوهش، بهبود نتایج استخراج ویژگی‌ها از متن نظرات فارسی و همچنین افزایش دقت تحلیل احساسات در این زمینه می‌باشد. از این رو با توجه به شکل ۱، چارچوب جدیدی برای تحلیل احساسات فارسی در سطح جمله و بر اساس مدل تعبیه‌سازی BERT، مدل ترکیبی استخراج ویژگی CNN-BiLSTM و همچنین الگوریتم XGBoost برای طبقه‌بندی احساسات در سه گروه مثبت، منفی و خنثی پیشنهاد شده است.

[1] Gan et al.
[2] Dashtipour et al.
[3] Multimodal sentiment analysis
[4] One-dimensional sentiments analysis

۳-۱. مجموعه داده

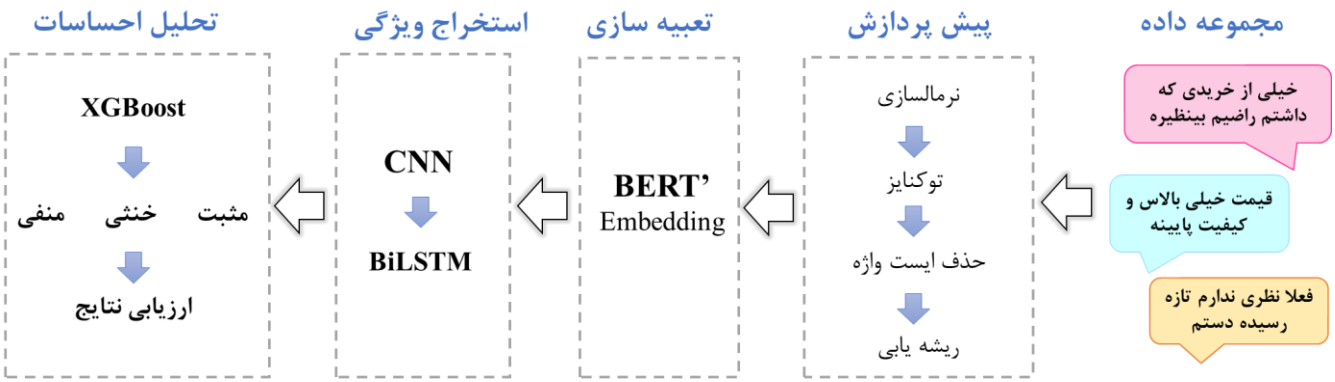
در دنیا بیش از ۱۲۰ میلیون نفر از زبان فارسی به عنوان زبان اصلی و زبان دوم خود استفاده می‌کنند [۲۰]. با این حال، ضمن افزایش محتوای تولید شده توسط کاربران فارسی زبان در انواع وب‌سایت‌ها و پلتفرم‌های آنلاین، همچنان عدم دسترسی به مجموعه داده مناسب متون فارسی، از جمله چالش‌های مهم حوزه تحلیل احساسات فارسی محسوب می‌شود. این در حالیست که در این پژوهش، مجموعه داده‌ای بر اساس متون نظرات کاربران فارسی زبان در رابطه با محصولات دیجیتالی وب‌سایت خرده‌فروشی آنلاین دیجی‌کالا تشکیل و در دیگر مراحل پژوهش مورد بررسی قرار گرفته است.

۳-۲. پیش پردازش

زبان فارسی به دلیل برخورداری از چندین شکل نوشتاری برای بعضی از حروف الفبا، انواع گویش‌ها[5] و لهجه‌ها، انواع فاصله[6] و نیم_فاصله[7] بین کلمات، به یکی از زبان‌های چالش برانگیز در حوزه تحلیل احساسات تبدیل شده است [۲۱]. این در حالیست که کیفیت و ساختار صحیح واژگان به واسطه اجرای دقیق مراحل پیش‌پردازش، از جمله مواردی است که باعث بهبود عملکرد طبقه‌بندی و تحلیل احساسات فارسی می‌شود [۲۲]. بنابراین مراحل به کار رفته برای پیش‌پردازش نظرات فارسی در این پژوهش عبارتند از:

- نرمال‌سازی[8]: به فرآیند تبدیل متن بدون ساختار[9] به فرمت استاندارد و یکپارچه بر اساس یکسان سازی تمامی حروف، تبدیل اعداد به کلمات معادل آنها و حذف علائم نگارشی؛ نرمال سازی می‌گویند [۲۳].
- توکن‌سازی[10]: در طی اجرای توکن‌سازی، پاراگراف‌های بزرگ متنی با توجه به علائم نگارشی موجود در متن (مانند نقطه، علامت تعجب، علامت سوال و ...)، به جملات، شکسته می‌شوند و سپس هر یک از جملات بر اساس فاصله بین کلمات، به نشانه‌ها و یا توکن‌های[11] معنادار تفکیک می‌شوند [۲۳].

[5] Dialectal varieties
[6] Space
[7] Half-space
[8] Normalization
[9] unstructured
[10] Tokenization
[11] Tokens



شکل ۱. چارچوب پیشنهادی پژوهش

- حذف ایست‌واژه [1]: هدف اصلی حذف ایست‌واژه، کنار گذاشتن کلماتی است که در تحلیل احساسات نقشی ندارند و تنها به دلیل تکرار زیاد در متون، باعث افزایش ابعاد مجموعه داده شده‌اند [۲۴]؛ بنابراین اجرای مرحله حذف ایست‌واژه، نه تنها باعث کاهش ابعاد مجموعه داده می‌شود؛ بلکه افزایش سرعت و ارائه نتایج دقیق‌تر مدل طبقه‌بندی را نیز منجر می‌شود [۲۴].
- ریشه‌یابی [2]: تکنیک ریشه‌یابی در تلاش است تا با هدف کاهش ابعاد مجموعه داده، زمان‌های دستوری کلمات را به شکل پایه آن‌ها تغییر دهد [۲۳].

۳-۳. تعبیه کلمات

تعبیه کلمات، روشی برای نگاشت داده‌های متنی به بردارهای عددی است. در این پژوهش برای تبدیل نظرات به بردارهای عددی از مدل BERT استفاده می‌شود؛ زیرا BERT به دلیل استفاده از یادگیری انتقالی [3] و مکانیزم توجه [4]، توانسته است زبان طبیعی را به واسطه فهم زمینه [5] و ارتباط بین کلمات، به شکل عمیق‌تری درک کند [۲۵].

۳-۴. استخراج ویژگی

ویژگی‌های نامرتبط موجود در متون نظرات، در طی فرآیند استخراج ویژگی، با هدف کاهش ابعاد و مدت زمان پردازش و همچنین افزایش دقت مدل طبقه‌بندی حذف می‌شوند [۲۶]. در این پژوهش از ترکیب دو مدل یادگیری عمیق CNN و Bi-LSTM، برای استخراج ویژگی‌ها استفاده می‌گردد. زیرا CNN از جمله روش‌های قدرتمند در زمینه استخراج ویژگی‌ها است، که می‌تواند ویژگی‌های محلی [6] بین کلمات متوالی یک جمله را به شکل دقیق‌تری استخراج نماید [۱۴-۱۵]. این در حالیست که CNN اطلاعات معنایی میان کلمات جمله یا سند را نادیده می‌گیرد و نمی‌تواند وابستگی‌های خیلی دور بین عبارات و جملات را یاد بگیرد [۵,۱۵]. بنابراین برای غلبه بر این محدودیت CNN، از مدل Bi-LSTM استفاده می‌گردد. زیرا Bi-LSTM ضمن پردازش داده‌ها در دو جهت، توالی و ترتیب زمانی بین کلمات موجود در یک سند را به شکل بهتری حفظ می‌کند و درعین حال برای استخراج ویژگی‌ها، زمینه‌ی متن [7] را نیز در نظر می‌گیرد [۴]. از این رو در این پژوهش، با بهره‌گیری از مدل ترکیبی "CNN-BiLSTM"، با استخراج حداکثر ویژگی‌ها توسط CNN و با در نظر گرفتن ترتیب زمانی بین کلمات و حذف کلمات غیرضروری توسط Bi-LSTM، ویژگی‌ها به شکل بهتری فیلتر می‌گردند؛ که همین مسئله باعث بهبود عملکرد مرحله استخراج ویژگی‌ها و در نهایت افزایش دقت مدل طبقه‌بندی احساسات می‌شود [۴].

۳-۵. تحلیل احساسات

"تحلیل احساسات"، به فرآیند استخراج نگرش‌های موجود در متون و طبقه‌بندی آن‌ها در یکی از گروه‌های منفی، مثبت و خنثی، گفته می‌شود [۲۸]. در این پژوهش برای تحلیل و طبقه‌بندی نگرش‌های مرتبط با محصولات دیجیتال، از رویکرد یادگیری ماشین و الگوریتم XGBoost استفاده می‌گردد؛ زیرا XGBoost به دلیل قابلیت محاسبات موازی [8]، سرعت پردازش بالا، قابلیت تنظیم پارامترها [9] بر اساس نوع مسئله و همچنین قدرت بالا در پردازش داده‌های نامتوازن [10]، از جمله الگوریتم‌های قدرتمند در زمینه تحلیل احساسات می‌باشد [۲۹].

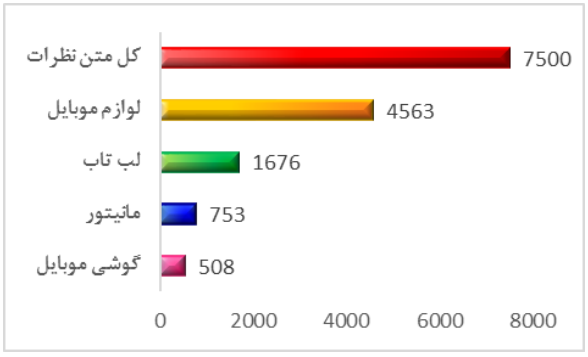
[1] Removing stop words
[2] Stemming
[3] Transfer Learning
[4] Attention Mechanism
[5] Context
[6] Local features
[7] Text Contexts
[8] Parallel Computing
[9] Parameters Tuning
[10] Unbalanced Data

۴. پیاده‌سازی و تحلیل نتایج

برای پیاده‌سازی هر مرحله از چارچوب پیشنهادی، از زبان برنامه‌نویسی پایتون در محیط گوگل کولب و تنظیمات GPU استفاده شده است.

۴-۱. مجموعه داده

در این پژوهش، بر اساس مجموعه نظرات فارسی دیجی‌کالا[1] که در سایت Kaggle موجود است، مجموعه داده‌ایی شامل ۷۵۰۰ نظر ارائه شده در زمینه محصولات دیجیتال وب‌سایت خرده‌فروشی آنلاین دیجی‌کالا، ایجاد شد. در طی بررسی اولیه مشخص گردید مجموعه داده، علاوه بر متون نظرات، شامل فیلدهای اطلاعاتی شناسه کاربر، شناسه محصول، عنوان محصول، پیشنهاد محصول و عدم پیشنهاد محصول می‌باشد. همچنین مشخص گردید، نظرات موجود در مجموعه داده در زمینه ۴ دسته‌بندی مختلف از محصولات دیجیتال می‌باشند (شکل ۲). علاوه بر این در طی این بررسی مشخص شد، طول نظرات ارائه شده، متفاوت می‌باشد؛ به گونه‌ای که از میان ۷۵۰۰ متن نظر، ۱۷۲۵ نظر یعنی ۲۳٪ از نظرات تنها بر اساس ۳ کلمه (توکن) و ۵۷۷۵ نظر یعنی ۷۷٪ از نظرات از بیش از ۲۰ کلمه (توکن) تشکیل شده‌اند. مانند نظر "این خیلی خوبه" که شامل ۳ کلمه و یا نظر "گوشی خوبیه بین این گوشی و چند گوشی دیگه شک داشتم اما از خریدم راضیم این گوشی با این کانفیگ و این قیمت ارزش خرید داره" که شامل ۲۶ کلمه (توکن) می‌باشند.



شکل ۲. تعداد نظرات مرتبط با دسته‌بندی محصولات

اما از نکات حائز اهمیت در این پژوهش، استفاده از فیلد اطلاعاتی "پیشنهاد محصول" و "عدم پیشنهاد محصول"؛ برای تعیین قطبیت[2] نظرات می‌باشد. بدین صورت که برای نظرات دارای گزینه "عدم پیشنهاد محصول" قطبیت منفی، برای نظرات دارای "پیشنهاد محصول" قطبیت مثبت و برای نظراتی که فاقد این گزینه و واژگان احساسی بودند، قطبیت خنثی در نظر گرفته شد. بنابراین مجموعه داده با توجه به برچسب‌های نسبت داده شده، نامتوازن و شامل ۳۲۰۰ نظر مثبت، ۲۸۰۰ نظر منفی و ۱۵۰۰ نظر خنثی می‌باشد.

۴-۲. پیش پردازش

برای اجرای مراحل مختلف پیش‌پردازش در این بخش از پژوهش، از کتابخانه Parsivar پایتون استفاده شده است. بدین صورت که در ابتدا مشخص گردید، متون نظرات، شامل مواردی همچون استفاده از حروف، کلمات و اعداد زبان‌های دیگر به ویژه زبان انگلیسی و عربی، عدم رعایت فاصله و نیم فاصله بین کلمات و عبارات می‌باشند. از این رو در ابتدا فرآیند نرمال‌سازی با استفاده از ماژول "Normalizer" بر روی نظرات اعمال گردید. سپس از ماژول "Tokenizer"، برای تفکیک و شکستن نظرات به جملات تشکیل دهنده بر اساس علائم نگارشی موجود در آن‌ها، استفاده شد. همچنین مجدداً از این ماژول برای شکستن هر یک از جملات، به کلمات تشکیل دهنده‌اش، استفاده گردید. اما نکته حائز اهمیت در پیش‌پردازش زبان فارسی، عدم وجود ماژولی در کتابخانه‌های موجود برای پیش‌پردازش این زبان، به منظور اجرای مرحله حذف ایست‌واژه می‌باشد. بنابراین در این پژوهش، برای اجرای این مرحله، تابعی با استفاده از زبان پایتون تحت عنوان remove-stop-words ایجاد گردید، که در آن شناسایی و حذف ایست‌واژه‌ها، بر اساس یک لیست ۱۳۱۶ ایست‌واژه فارسی [3] صورت گرفته است. همچنین در نهایت ماژول "FindStems" به‌عنوان آخرین مرحله‌ی پیش‌پردازش و برای اجرای فرآیند ریشه‌یابی بر روی مجموعه داده اعمال گردید.

۴-۳. تعبیه کلمات

برای تبدیل اطلاعات متنی موجود در نظرات به بردارهای عددی، از ParsBERT الگوریتم BERT استفاده شده و حداکثر طول ورودی برای آن در این پژوهش، ۵۱۲ توکن و Embedding dimension برابر با ۷۶۸ درنظر گرفته شده است.

۴-۴. استخراج ویژگی

در این پژوهش، استخراج ویژگی‌ها بر اساس دو مرحله "استخراج ویژگی‌های محلی از بردارهای عددی متون نظرات"، "استخراج ویژگی‌های مرتبط با زمینه متن" و با استفاده از مدل

[1] https://www.kaggle.com/datasets/soheiltehranipour/digikala-comments-persian-sentiment-analysis
[2] Polarity
[3] https://github.com/kharazi/persianstopwords/blob/master/persian

ترکیبی CNN-BiLSTM انجام شده که، روش کار بدین صورت است:

در ابتدا و برای اجرای مرحله اول، خروجی مدل BERT (بردارهای عددی متون نظرات)، به مدل CNN ارائه شده، که در ادامه، CNN فرآیند استخراج ویژگی‌های محلی را با استفاده از لایه‌های Convolution، Max-pooling و Dropout، بر روی آن‌ها اعمال کرده است. سپس برای اجرای مرحله دوم، یعنی استخراج ویژگی‌های مرتبط با زمینه متن، خروجی CNN به عنوان ورودی به مدل BiLSTM داده شده است. سپس BiLSTM با توجه به زمینه متن، توالی زمانی بین کلمات و حذف کلمات غیر ضروری، استخراج ویژگی‌های مرتبط با زمینه متن را انجام داده است. لازم به ذکر است که، CNN و BiLSTM برای پردازش متن دارای پارامترهای مشخصی هستند که به طور معمول و در تمامی پژوهش‌ها از این پارامترها برای پردازش متن، استفاده می‌گردد. از این رو در جدول ۱، به ساختار و مقادیر به کار رفته برای هر یک از این پارامترها در این بخش از پژوهش و با هدف ایجاد مدل CNN-BiLSTM، اشاره شده است.

جدول ۱. ساختار CNN_BiLSTM

لایه ها	مقدار پارامتر هر لایه
Convolution1D	Filters= ۱۲۸ kernel_size =۳ activation = ‘relu’
MaxPooling Dropout Convolution1D	pool_size= ۲ Rate = ۰/۵ Filters= ۶۴ kernel_size =۳ , activation = ‘relu’
MaxPooling Dropout Convolution1D	pool_size=۲ Rate = ۰/۵ Filters= ۳۲ kernel_size =۳ activation = ‘relu’
MaxPooling Dropout Bidirectional(LSTM)	pool_size=۲ Rate =۰/۲ Units = ۲۵۰

با این حال برای اجرای مدل CNN-BiLSTM، لازم است تا مقادیر پارامترهای Optimizer، Loss Function، Epochs و Batch size نیز تعیین گردد. بنابراین برای تنظیم پارامترهای[1] اشاره شده از تکنیک Grid Search استفاده شد؛ زیرا مقدار بهینه برای هر یک از این پارامترها باید به گونه‌ای در نظر گرفته شود تا مدل ایجاد شده بهترین دقت را به عنوان نتیجه نهایی ارائه دهد. جدول ۲ به مقادیر بهینه به دست آمده، برای هر یک از پارامترهای مطرح شده، اشاره دارد.

جدول۲. مقدار بهینه پارامترهای مدل ترکیبی CNN-BiLSTM

پارامترها	مقدار پارامترها
Optimizer	'adam'
Loss Function	categorical_crossentropy
Epochs	۷۰
Batch size	۱۰۰

۵-۴. طبقه‌بندی و تحلیل احساسات

در این مرحله برای طبقه‌بندی احساسات در سطح جمله از الگوریتم XGBoost استفاده می‌گردد. از این رو با هدف ایجاد مدل طبقه‌بندی ۶۰٪ مجموعه داده برای آموزش (۴۵۰۰ متن نظر)، ۳۰٪ برای آزمایش (۲۲۵۰ متن نظر) و ۱۰٪ برای اعتبارسنجی (۷۵۰ متن نظر) در نظر گرفته شده است. لازم به ذکر است، XGBoost پارامترهای فراوانی دارد که می‌توان آنها را بر اساس نوع مسئله و با هدف بهبود دقت مدل طبقه‌بندی تنظیم نمود. بنابراین برای تنظیم پارامترهای اثرگذار XGBoost در دقت طبقه‌بندی و تحلیل احساسات، از روش Grid Search بر روی داده‌های اعتبارسنجی استفاده گردید که جدول ۳ به مقادیر بهینه‌ی به دست آمده برای هر یک از این پارامترها اشاره دارد.

جدول ۳. مقدار بهینه پارامترهای XGBoost

پارامترها	مقادیر پارامترها
learning_rate	۰/۱
Booster	"gbtree"
n_estimator	۱۰۰
Objective	’multi:softprob’
max_depth	۸
min_child_weight	۶
Subsample	۰/۷

در ادامه مدل طبقه‌بندی XGBoost بر اساس مقادیر بهینه‌ی پارامترها، ایجاد و بر روی مجموعه داده اعمال گردید. نتایج نشان می‌دهد XGBoost، خروجی مراحل BERT و CNN-BiLSTM را با دقت ۹۵/۰۲٪ و در سه گروه منفی، مثبت و خنثی طبقه‌بندی کرده است. سپس برای اطمینان از صحت دقت به دست آمده، عملکرد مدل بر اساس ماتریس

[1] Parameters tuning

درهم‌ریختگی[1] و معیارهای Accuracy، Precision، Recall، F1–Score مورد بررسی قرار گرفت. همچنین برای اطمینان از درستی و صحت دقت به دست آمده، از یکی دیگر از روش‌های ارزیابی عملکرد الگوریتم‌های یادگیری ماشین با عنوان K-Fold نیز استفاده گردید که مقدار K برای آن برابر با ۵ در نظر گرفته شده است. در جدول ۴ به نتایج به دست آمده برای هر یک از این معیارها اشاره شده است.

جدول ۴. ارزیابی عملکرد الگوریتم XGBoost (بر اساس داده‌های آزمایشی)

Confusion Matrix		مقادیر پیش‌بینی شده			تعداد نظرات هر دسته
		برچسب +۱	برچسب ۰	برچسب -۱	
مقادیر واقعی	برچسب +۱	۹۷۹	۱۲۷	۸۱	۱۱۸۷
	برچسب ۰	۳۴	۸۹	۱۳	۱۳۶
	برچسب -۱	۹۴	۳۴	۷۹۹	۹۲۷

معیارهای ارزیابی	Accuracy	Precision	Recall	f1-score
نتایج	۹۵/۰۲	۹۴	۹۴	۹۴
ارزیابی بر اساس 5-fold	۹۴/۱۲	۹۳/۶۸	۹۴	۹۳/۸۳

۶-۴. مقایسه نتایج

در این پژوهش، مقایسه نتایج بر اساس سه مرحله "مقایسه و ارزیابی عملکرد مدل‌های استخراج ویژگی CNN-BiLSTM و CNN-LSTM"، "مقایسه روش پیشنهادی بر روی مجموعه داده‌های مختلف در زبان فارسی" و "مقایسه نتایج چارچوب پیشنهادی با دیگر کارهای صورت گرفته در حوزه تحلیل احساسات فارسی"، انجام شده است.

۱-۶-۴. مقایسه CNN-BiLSTM و CNN-LSTM

در این مرحله، برای ارزیابی تاثیر عملکرد مدل استخراج ویژگی، بر روی دقت مدل طبقه‌بندی در تحلیل احساسات فارسی، به مقایسه عملکرد مدل‌های استخراج ویژگی CNN-BiLSTM و CNN-LSTM بر اساس مجموعه داده به کار رفته، پرداخته شده است. اما پیش از این مقایسه، مدل استخراج ویژگی CNN-LSTM بر اساس ساختار اشاره شده در جدول ۵ بر روی بردارهای عددی متون نظرات (خروجی حاصل از مرحله BERT) اعمال گردید. سپس مدل XGBoost برای طبقه‌بندی احساسات در سه گروه مثبت و منفی و خنثی، بر روی ویژگی‌های استخراج شده اجرا گردید که نتایج دقت ۹۳٪/۵۶ را نشان می‌دهد.

جدول ۵. ساختار CNN-LSTM

لایه ها	مقدار پارامترهای هر لایه
Convolution1D	Filters=۱۲۸ kernel_size= ۳ activation = ‘relu’
MaxPooling	pool_size=۲
Dropout	Rate = ۰/۵
Convolution1D	Filters=۶۴ kernel_size= ۳ ,activation = ‘relu’
MaxPooling	pool_size=۲
Dropout	Rate = ۰/۵
Convolution1D	Filters=۳۲ kernel_size= ۳ activation = ‘relu’
MaxPooling	pool_size=۲
Dropout	Rate = ۰/۵
LSTM	Units =۲۰۰
پارامترها	**مقدار پارامترها**
Optimizer	'adam'
Loss Function	categorical_crossentropy
Epochs	۷۰
Batch size	۲۰۰

در نهایت باتوجه به جدول ۶، مشخص گردید که XGBoost، خروجی مدل استخراج ویژگی CNN-BiLSTM را با دقت ۹۵٪/۰۲ و F1-score = ٪۹۴ و خروجی مدل استخراج ویژگی CNN-LSTM را با دقت ۹۳/۵۶٪ و مقدار F1-score برابر با ۹۲/۵۱٪ طبقه‌بندی کرده است. از این رو با توجه به برتری دقت به دست آمده برای مدل طبقه‌بندی XGBoost بر اساس CNN-BiLSTM، می‌توان اذعان داشت که CNN-BiLSTM در استخراج ویژگی‌ها از متون نظرات فارسی به واسطه استفاده از BiLSTM و پردازش داده‌ها در دو جهت، دارای عملکرد بهتری نسبت به LSTM در مدل CNN-LSTM می‌باشد. به‌گونه‌ای که ضمن بهبود عملکرد و نتایج، در

[2] Confusion Matrix

مرحله استخراج ویژگی‌ها از متون نظرات فارسی، افزایش دقت در مرحله طبقه‌بندی احساسات را نیز به دنبال داشته است.

جدول ۶. مقایسه CNN-BiLSTM و CNN-LSTM

	مدل طبقه‌بندی				معیارهای ارزیابی			
	Word Embedding	Feature Extraction	Classifier	Accuracy	Precision	Recall	F1-score
روش پیشنهادی برای مقایسه	BERT	CNN-LSTM	XGBoost	۹۳.۵۶	**۹۳.۰۲**	**۹۲**	**۹۲.۵۱**
روش پیشنهادی پژوهش	**BERT**	**CNN-BiLSTM**	**XGBoost**	**۹۵.۰۲**	**۹۴**	**۹۴**	**۹۴**

۲-۶-۴. مقایسه نتایج روش پیشنهادی بر روی مجموعه داده‌های مختلف در زبان فارسی

در این بخش، به مقایسه و تحلیل عملکرد روش پیشنهادی در این پژوهش، بر روی دیگر مجموعه داده‌های معتبر در زبان فارسی پرداخته شده است.

- مجموعه داده نظرات هتل[1]:

مجموعه داده شامل ۶۴۳۳ متن نظر فارسی می‌باشد که با حذف نظراتی مانند "سلام"، "خداحافظ دوستان" و نظرات تکراری؛ که تاثیری در تجزیه و تحلیل احساسات ندارند، تعداد آن‌ها به ۶۱۲۵ متن نظر رسید. طول نظرات در این مجموعه داده متفاوت و کوتاه‌ترین نظر از ۳ کلمه "بسیار راضی هستیم" و طولانی‌ترین از ۸۰ کلمه تشکیل شده است مانند "واسه ماه عسل رفتیم منم با نظر بقیه موافقم به جز دور بودن مسیر و کرایه رفت و برگشت به همه جا که خیلی گرون بود خوده هتل از نظر بهداشتی خوب بود از نظر برخورد و کیفیت رستوران هم بد نبود ولی مشکل اصلی این بود که سر ساعت خاصی از نهار خبری نبود و این خیلی باعث بینظمی برنامه های ما میشد ولی در هر صورت به من که خیلی خیلی خوش گذشت و پیشنهاد میکنم به بقیه". همچنین مجموعه داده نامتوازن و شامل ۳۴۴۹ نظر مثبت، ۲۱۷۲ نظر منفی و ۵۰۴ نظر خنثی می‌باشد.

- مجموعه داده نظرات پیرامون تلفن همراه[2]:

این مجموعه داده از ۶۵۶۸ متن نظر تشکیل گردیده که کوتاه‌ترین آن شامل ۲ کلمه "کاملترین گوشی" و طولانی‌ترین آن شامل ۱۱۷ کلمه مانند "حدود یک ماهی میشه این گوشی رو گرفتم و مشکل داغ شدن ندارم فقط در بازی سنگین پشتش گرم میشه که برای همه گوشیها عادیه کسایی هم که از این گوشی بد میگن به من ربطی نداره چون من همه نظرات رو خوندم و بازم گرفتمش واسه اینکه تا با گوشی کار نکنید نمیتونین در موردش تصمیمی بگیرین و نظر بدین در ضمن دوستانی که قصد خرید این گوشی رو دارن فقط گوشیی رو بگیرن که مونتاژ ۲۰۱۴ باشه چون بهترین بهینه سازی روی اون انجام شده و حداکثره بهره‌وری رو ارائه میده پس در کل این گوشی عالیه اگه کره ای یا ویتنامیش گیرتون اومد و مونتاژ ۲۰۱۴ بود در خریدش شک نکنید و حتما بخرید" می‌باشد. لازم به ذکر است که این مجموعه داده نامتوازن و از ۲۹۸۱ نظر مثبت، ۲۶۶۳ نظر منفی و ۹۲۴ نظر خنثی می‌باشد.

در ادامه هر یک از مجموعه داده‌ها به سه دسته آموزش (۶۰٪)، آزمایش (۳۰٪) و اعتبارسنجی (۱۰٪) تقسیم شدند. سپس روش Grid Search با هدف تنظیم پارامترها بر روی داده های اعتبارسنجی و مدل پیشنهادی در این پژوهش (شامل استفاده از BERT برای تبدیل نظرات به بردارهای عددی، CNN-BiLSTM برای استخراج ویژگی‌ها و XGBoost برای طبقه‌بندی احساسات) با هدف طبقه‌بندی احساسات موجود در متون نظرات به سه گروه منفی و خنثی و مثبت، به صورت جداگانه بر روی هر یک از مجموعه داده‌ها پیاده‌سازی شدند. در نهایت و با توجه به جدول ۷، مشخص گردید که، روش پیشنهادی پژوهش توانسته است بر روی مجموعه داده اولیه (نظرات محصولات دیجیتال دیجی‌کالا) که دارای ۷۵۰۰ نظر با طول‌های متفاوت بین ۳ تا ۲۶ کلمه می‌باشد، به دقت ۹۵/۰۲ درصد دست یابد. درحالیکه همین روش بر روی مجموعه داده

[1] https://hooshbox.ir
[2] https://hooshbox.ir

جدول ۷. مقایسه نتایج روش پیشنهادی بر روی مجموعه داده‌های فارسی

مجموعه نظرات	اطلاعات مجموعه داده			معیارهای ارزیابی			
	تعداد نظرات	طول کوتاه‌ترین متن نظر	طول بلند ترین متن نظر	Accuracy	Precision	Recall	F1-score
محصولات دیجیتال دیجی‌کالا	۷۵۰۰	۳	۲۶	۹۵.۰۲	۹۴	۹۴	۹۴
نظرات هتل	۶۲۱۵	۳	۸۰	۸۸.۳۱	۸۷.۰۲	۸۶	۸۶.۵۰
نظرات تلفن همراه	۶۵۶۸	۲	۱۱۷	۸۱.۴۴	۸۱	۸۰	۸۰.۴۹

هتل که به نسبت مجموعه داده اصلی دارای نظراتی با طول‌های متفاوت‌تر (بین ۳ تا ۸۰ کلمه) می‌باشد، به دقت ۸۸.۳۱ درصد رسیده است. اما نکته حائز اهمیت در آن است که همین روش بر روی مجموعه داده تلفن همراه، که به نسبت دو مجموعه داده قبلی شامل نظراتی با طول‌های بسیار متفاوت‌تر (بین ۲ تا ۱۱۷ کلمه) می‌باشد؛ به دقت ۸۱.۴۴ درصد دست یافته است. بنابراین بر اساس نتایج به دست آمده در این بخش می‌توان اینگونه اذعان داشت که طول متون نظرات از جمله عوامل اثرگذار بر روی کارایی و دقت مدل طبقه‌بندی و تحلیل احساسات می‌باشد. به گونه‌ایی که وجود متون نظرات با طول مشابه و نزدیک به هم در یک مجموعه داده باعث افزایش کارایی و دقت مدل طبقه‌بندی احساسات می‌شوند (مانند مجموعه داده نظرات دیجی‌کالا)، درحالیکه وجود متون نظرات با طول‌های بسیار متفاوت در یک مجموعه داده، نه تنها به مراحل پیچیده‌ایی از پردازش نیازمند است بلکه می‌تواند باعث کاهش کارایی و دقت مدل طبقه‌بندی و تحلیل احساسات نیز گردند (مانند مجموعه داده نظرات تلفن همراه و پس از آن مجموعه نظرات هتل).

۳-۶-۴. مقایسه نتایج پژوهش با دیگر کارهای تحلیل احساسات نظرات فارسی

در این بخش به مقایسه و تحلیل عملکرد روش پیشنهادی، با دیگر پژوهش‌های فارسی که بر روی مجموعه داده دیجی‌کالا صورت پذیرفته‌اند، پرداخته شده است. لازم به ذکر است، مبنای مقایسه در این بخش چگونگی استفاده از مدل‌های CNN و BiLSTM (استفاده به صورت مجزا و یا ترکیب مدل‌ها) در مراحل مختلف تحلیل احساسات و بررسی تأثیر آن بر روی عملکرد نهایی می‌باشد. با توجه به پژوهش‌های اشاره شده در جدول ۸، شرامی و همکاران [۱۲] و همچنین شمالی و همکاران [۱۳] اذعان داشته‌اند، افزایش دقت تحلیل احساسات زبان فارسی در صورتی امکان‌پذیر می‌باشد که از الگوریتم‌های پر قدرتی مانند CNN و BiLSTM برای طبقه‌بندی احساسات استفاده شود. این در حالیست که دشتی‌پور و همکاران [۱۹] تنها از مدل BiLSTM برای استخراج ویژگی‌ها استفاده کرده‌اند و معتقدند، این روش با در نظر گرفتن وابستگی‌های طولانی مدت میان متن نظرات، می‌تواند باعث بهبود عملکرد مدل طبقه‌بندی شود. اما در این پژوهش، ادعای اصلی مطرح شده آن است که، زمانی می‌توان به نتایج دقیقی در تحلیل و طبقه‌بندی احساسات زبان فارسی دست یافت که، مرحله استخراج ویژگی‌ها به شکل دقیقی بر روی نظرات فارسی پیاده‌سازی شده باشد. بر همین اساس در این پژوهش، از ترکیب مدل‌های CNN و BiLSTM برای استخراج ویژگی‌ها استفاده شده است؛ زیرا CNN با تمرکز بر روی ویژگی‌های محلی، می‌تواند ویژگی‌های احساسی موجود در هر یک از متون نظرات فارسی را با دقت بیشتری پردازش و استخراج نماید. از سوی دیگر با توجه به اینکه در زبان فارسی، نظرات عموماً طولانی و دارای ویژگی‌های احساسی مختلف و متعدد در طول متن نظر می‌باشند؛ مدل BiLSTM به واسطه در نظر گرفتن وابستگی‌های طولانی مدت میان متون و همچنین یاد گرفتن زمینه متن، می‌تواند ویژگی‌های احساسی و مرتبط با زمینه متون نظرات را به شکل دقیق‌تری استخراج کند. با این حال در ادامه‌ی این پژوهش و برای اثبات این ادعا، عملکرد CNN-BiLSTM با CNN-LSTM مورد بررسی و مقایسه قرار داده شد؛ که نتایج نهایی نشان می‌دهد، BiLSTM به دلیل پردازش

داده‌ها در دو جهت، نسبت به LSTM که داده‌ها را به صورت یک طرفه پردازش می‌کند؛ موجب گردیده تا مدل CNN-BiLSTM به شکل دقیق‌تری استخراج ویژگی‌های متون نظرات فارسی را انجام دهد و در نهایت باعث بهبود و افزایش

جدول۸. مقایسه و تحلیل عملکرد چارچوب پیشنهادی پژوهش و روش‌ها و نتایج دیگر پژوهش‌های تحلیل احساسات فارسی

پژوهش	متد پژوهش				نتایج			
	مجموعه داده	Word embedding	استخراج ویژگی	مدل طبقه بندی	Accuracy		f1-score	
					۲ کلاس	۳ کلاس	۲ کلاس	۳ کلاس
شرامی و همکاران [۱۲]	متن نظرات دیجی‌کالا	Keras Embedding		CNN			۹۱/۹۰	۶۶/۶۵
				BiLSTM			۹۱/۹۸	۶۶/۵۰
		FastText		CNN			۸۸/۰۶	۶۶/۳۰
				BiLSTM			۹۰/۵۹	۶۹/۳۳
دشتی‌پور همکاران [۱۹]	متن نظرات دیجی‌کالا	FastText	BiLSTM	LSTM + Dependency-based		۸۸/۰۱		
شمالی و همکاران [۱۳]	متن نظرات دیجی‌کالا	FastText		CNN	۹۹/۶۰		۹۵/۶۰	
				BiLSTM	۹۹/۳۳		۹۲/۳۰	
چارچوب پیشنهادی	متن نظرات دیجی‌کالا	**BERT**	**CNN-BiLSTM**	**XGBoost**	**۹۵/۰۲**		**۹۴**	
		BERT	**CNN-LSTM**	**XGBoost**	**۹۳/۵۶**		**۹۲/۵۱**	

دقت مدل طبقه‌بندی XGBoost با نتایج، دقت ۹۵/۰۲٪ و F1-score=۹۴٪ شود.

۵. نتیجه‌گیری

در حال حاضر بیشتر تعاملات مشتریان و خرده‌فروشان بر اساس صفحات وب و دیگر پلتفرم‌های آنلاین صورت می‌گیرد. مشتریان محصولات موجود در فروشگاه‌های خرده‌فروشی آنلاین را بررسی و بعد از خرید، نظرات و تجربیات خود در زمینه محصول را با دیگران به اشتراک می‌گذارند. از این رو مدیران فروشگاه‌های خرده‌فروشی آنلاین در تلاشند تا با استفاده از تکنیک‌های تحلیل احساسات به خلاصه‌سازی و تحلیل نظرات منتشر شده بپردازند و بر اساس نتایج آن‌ها سیاست‌های جدیدی را برای تولید و ارتقا محصولات، افزایش فروش، جذب بیشتر مشتریان و مدیریت هرچه بهتر برند خود در نظر بگیرند. در این پژوهش، یک چارچوب جدید برای تحلیل احساسات نظرات فارسی در سطح جمله پیشنهاد شده است، که بر اساس مراحل آن، بعد از پیش‌پردازش نظرات، روش تعبیه‌سازی BERT بر روی مجموعه داده اعمال شده است.

سپس، مدل استخراج ویژگی CNN-BiLSTM بر روی متون نظرات و با هدف استخراج ویژگی‌های محلی و همچنین ویژگی‌های احساسی مرتبط با متن زمینه، پیاده‌سازی گردیده است. همچنین در ادامه مدل طبقه‌بندی XGBoost برای طبقه‌بندی احساسات به کار رفته است که نتایج دقت ۹۵/۰۲٪ و F1-score=۹۴٪ را برای این مدل نشان می‌دهد. با این حال، برای اطمینان از عملکرد صحیح CNN-BiLSTM در زمینه استخراج ویژگی‌ها از نظرات فارسی، به مقایسه CNN-BiLSTM و CNN-LSTM پرداخته شده است. نتایج نهایی نشان می‌دهد مدل طبقه‌بندی بر اساس CNN-BiLSTM به دقت ۹۵/۰۲٪ و F1-score= ۹۴٪ و بر اساس CNN-LSTM به دقت ۹۳/۵۶٪ و F1-score= ۹۲/۵۱٪ دست یافته است. بنابراین بر اساس نتایج، می‌توان اینگونه اذعان داشت، CNN-BiLSTM به واسطه استفاده از BiLSTM و پردازش دو طرفه

متون، روشی قدرتمند در زمینه استخراج ویژگی‌ها از متن نظرات فارسی است؛ که ضمن استخراج دقیق ویژگی‌ها، بهبود و افزایش دقت مدل طبقه‌بندی احساسات فارسی را به دنبال دارد. اما در کنار محدودیت‌های موجود برای ابزارهای پردازشی در زبان فارسی، می‌توان به کمبود مجموعه داده متنی معتبر، در زبان فارسی اشاره نمود که در صورت وجود آن‌ها، امکان پیاده‌سازی روش پیشنهادی و مقایسه نتایج بر روی مجموعه داده‌های مختلف، برای این پژوهش فراهم می‌گردید. با این حال و با وجود محدودیت‌های اشاره شده، در کارهای آتی، قصد داریم تا به بررسی عملکرد روش‌های خودرمزگذار[1] در زمینه استخراج ویژگی‌ها از متن نظرات فارسی و تاثیر آن بر روی دقت طبقه‌بندی و تحلیل احساسات فارسی، بپردازیم.

مراجع

[1] "2020-21 Annual Report | Digikala newsroom." https://www.digikala.com/mag/lp/report 2020-21/ (accessed Apr. 13, 2022).

[2] S. T K and J. Shetty, "Sentiment Analysis of Product Reviews," Encycl. Data Warehous. Mining, 2nd Ed., no. Icicct, pp. 1794–1799, 2017, doi: 10.5281/zenodo.264151.

[3] Q. Shen, Z. Wang, and Y. Sun, "Sentiment analysis of movie reviews based on machine learning," Int. Conf. Intell. Sci. (pp. 164-171). Springer, Cham., 2017, doi: 10.1145/3447490.3447491.

[4] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," Mach. Learn. Knowl. Extr., vol. 1, no. 3, pp. 832–847, 2019, doi: 10.3390/make1030048.

[5] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," IEEE Access, vol. 8, pp. 58080–58093, 2020, doi: 10.1109/ACCESS.2020.2982538.

[6 W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews : Sentiment analysis using lexicon integrated two-channel CNN – LSTM family models," Appl. Soft Comput. J., vol. 94, p. 106435, 2020, doi: 10.1016/j.asoc.2020.106435.

[7] N. J. Prottasha et al., "Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning," Sensors, vol. 22, no. 11, pp. 1–19, 2022, doi: 10.3390/s22114157.

[8] E. F. Ayetiran, "Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks," Knowledge-Based Syst., vol. 252, 2022, doi: 10.1016/j.knosys.2022.109409.

[9] H. Giradharbhai Nathani, Vivek Ketan Patel, "Twitter Sentiment Analysis for Hate speech detection using a C-BiLstm model," JAC A J. Compos. Theory, no. 0731–6755, pp. 1–13, 2022.

[10] K. Deepika, B. Sivanagaraju, A. Saikumar, E. B. V. N. P. Kumar, and B. Rohith, "Emotion Classification from Poetry text by using CNN-BiLSTM," JAC A J. Compos. Theory, no. 0731–6755, 2022.

[11] A. Yue, C. Mao, L. Chen, Z. Liu, C. Zhang, and Z. Li, "Detecting Changes in Perceptions towards Smart City on Chinese Social Media : A Text Mining and Sentiment Analysis," 2022.

[12] J. P. R. Sharami, P. A. Sarabestani, and S. A. Mirroshandel, "DeepSentiPers: Novel Deep Learning Models Trained Over Proposed Augmented Persian Sentiment Corpus," 2020, [Online]. Available: http://arxiv.org/abs/2004.05328.

[13] S. Shumaly, M. Yazdinejad, and Y. Guo, "Persian sentiment analysis of an online store independent of pre-processing using convolutional neural network with fastText embeddings," PeerJ Comput. Sci., vol. 7, pp. 1–22, 2021, doi: 10.7717/peerj-cs.422.

[14] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," Entropy, vol. 23, no. 5, pp. 1–16, 2021.

[15] ح. منصور، س. ممتازی و ک. لایقی, "تحلیل احساسات برای پیش‌بینی بازار بورس با شبکه عصبی ژرف: مطالعه موردی برای پایگاه داده سهام شرکت‌های بین-المللی," دوفصلنامه فناوری اطلاعات و ارتباطات ایران، سال سیزدهم، شماره های ۴۷ و ۴۸، بهار و تابستان, ۱۴۰۰.

[16] J. Yoon and H. Kim, "Multi-Channel Lexicon Integrated CNN-BiLSTM Models for Sentiment Analysis," Proc. 29th Conf. Comput. Linguist. speech Process. (ROCLING 2017) (pp. 244-253)., pp. 244–253, 2017.

[17] W. MENG, Y. WEI, P. LIU, Z. ZHU, and H. YIN, "Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM," IEEE

[1] autoencoder

Access, vol. 7, pp. 167240–167249, 2019, doi: 10.1109/ACCESS.2019.2952888.
[18] C. Gan, Q. Feng, and Z. Zhang, "Scalable multi-channel dilated CNN – BiLSTM model with attention mechanism for Chinese textual sentiment analysis," Futur. Gener. Comput. Syst., vol. 118, pp. 297–309, 2021, doi: 10.1016/j.future.2021.01.024.
[19] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "A novel context-aware multimodal framework for persian sentiment analysis," Neurocomputing, vol. 457, pp. 377–388, 2021, doi: 10.1016/j.neucom.2021.02.020.
[20] "All about Persian Language (Farsi) | ASPIRANTUM." https://aspirantum.com/blog/all-you-need-to-know-about-persian-language (accessed Apr. 13, 2022).
[21] S. Sarmady and E. Rahmani, "Word representation or word embedding in Persian texts," arxiv, 2017.
[22] E. Asgarian, M. Kahani, and S. Sharifi, "The Impact of Sentiment Features on the Sentiment Polarity Classification in Persian Reviews," Cognit. Comput., vol. 10, no. 1, pp. 117–135, 2018, doi: 10.1007/s12559-017-9513-1.
[23] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," Procedia Comput. Sci., vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.
[24] H. Saif, M. Fernandez, Y. He, and H. Alani, "On stopwords, filtering and data sparsity for sentiment analysis of twitter," Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014, pp. 810–817, 2014.
[25] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, no. Mlm, pp. 4171–4186, 2019.
[26] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," Eurasip J. Wirel. Commun. Netw., vol. 2017, no. 1, pp. 1–12, 2017, doi: 10.1186/s13638-017-0993-1.
[27] Y. Cheng, L. Yao, G. Xiang, G. Zhang, T. Tang, and L. Zhong, "Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU with Attention Mechanism," IEEE Access, vol. 8, pp. 134964–134975, 2020, doi: 10.1109/ACCESS.2020.3005823.
[28] B. Liu, "Sentiment Analysis and Opinion Mining," Lang. Arts Discip., no. May, p. 167, 2012.
[29] F. I. Rustam et al., "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," 2021, doi: 10.1371/journal.pone.0245909.