تحلیل احساسات رمزارزها با یادگیری انتقالی شات صفر

کیا جهان‌بین[1]، کاندید دکتری مهندسی نرم‌افزار، محمد علی زارع چاهوکی[2*]، دانشیار

[1 و 2] گروه کامپیوتر و نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران [1]kia.jahanbin@stu.yazd.ac.ir,
[2*]chahooki@yazd.ac.ir

چکیده: تحلیل احساسات جنبه‌محور توئیت‌ها با شبکه‌های عصبی عمیق بسیار مورد توجه است. با این حال، این شبکه‌ها برای کارایی مطلوب نیازمند دادگان غنی می‌باشند. بنابراین، در این مطالعه از یادگیری انتقالی عمیق شات صفر و دادگان میان‌دامنه‌ای برای افزایش غنای منابع آموزشی در دامنه هدف استفاده شده است. در این پژوهش، ابتدا توئیت‌های تاثیرگذاران رمزارزها جمع‌آوری شده و سپس، از ترکیب شبکه عصبی عمیق، لایه توجه و شبکه پیش‌آموزش‌دیده دستیل‌برت برای تحلیل احساسات جنبه‌محور استفاده شده است. نقشه حرارتی لایه توجه نشان می‌دهد استفاده از این لایه بعد از مدل‌های عمیق توانسته به برجسته‌سازی کلمات جنبه کمک کند. در یادگیری انتقالی شات صفر، مدل هیچ نمونه‌ای از داده‌های هدف دارای برچسب را در اختیار ندارد و آموزش با دادگان میان دامنه‌ای انجام می‌شود. علاوه بر این، استفاده از دادگان میان‌دامنه‌ای و انتخاب بهینه آنها با از ضریب شباهت پیرسون در شرایط کمبود داده غنی سبب کاهش انتقال منفی شده و همچنین حساسیت مدل را به محتوی زمینه متن کمک می‌کند. همچنین، برای حل مشکل دادگان نامتوازن از رویکرد نمونه‌برداری بیش از حد اقلیت مصنوعی استفاده شده است. یافته‌های تجربی نشان می‌دهد، مدل در مقایسه با پژوهش‌های پیشین، به صورت میانگین دقت و F1 را بر روی مجموعه داده بنچ‌مارک SemEval در حدود ۲ درصد افزایش داده و میانگین نرخ ROC-AUC به ۸۶,۳۵ درصد رسیده است. همچنین، در این پژوهش از نرخ یادگیری وفق‌پذیر استفاده شده که سبب همگرایی سریع‌تر مدل می‌شود. انتخاب ابرپارامترهای شبکه نیز با روش گریدی صورت گرفته که موجب وفق‌پذیری بیشتر مدل با دامنه هدف شده است.

واژه‌های کلیدی: تحلیل احساسات جنبه‌محور رمزارز، یادگیری عمیق شات صفر، شبکه‌های عصبی عمیق پیش‌آموزش دیده، لایه توجه، روش گریدی، نرخ یادگیری وفق‌پذیر.

* محمدعلی زارع چاهوکی، **chahooki@yazd.ac.ir**

Sentiment analysis of cryptocurrencies with Zero-shot transfer learning

Kia Jahanbin [1], PhD candidate Computer Engineering, Mohammad Ali Zare Chahooki [2*], Associate Professor

[1, 2] Department of Computer Engineering, Yazd University, Yazd, Iran, [1]kia.jahanbin@stu.yazd.ac.ir, [2*]chahooki@yazd.ac.ir

Abstract: Aspect-based sentiment analysis (ABSA) of tweets using deep neural networks has garnered significant attention. However, these networks generally perform best with access to extensive, high-quality data. This research leverages zero-shot deep transfer learning and cross-domain data to enrich the educational resources available within the target domain. The study began by gathering tweets from cryptocurrency influencers, followed by employing a combination of a deep neural network, an attention mechanism, and the pre-trained DistillBERT model for aspect-based sentiment analysis. The attention layer's heat map indicates that integrating this layer with deep models enhances the identification of aspect-specific words. In zero-shot transfer learning, the model operates without labelled target data samples, relying exclusively on cross-domain data for training. Additionally, using cross-domain data with optimal selection based on Pearson's similarity coefficient helps minimize negative transfer under low-data conditions, improving the model's sensitivity to textual context. To address data imbalance, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied. Experimental results indicate a 2% improvement in average accuracy and F1 score on the SemEval benchmark dataset, with the ROC-AUC reaching 86.35% relative to previous studies. Furthermore, we employed an adaptive learning rate to expedite model convergence and used Grid Search to optimize the network's hyperparameters, enhancing its adaptability to the target domain.

Keywords: *Cryptocurrency; Aspect-briented sentiment analysis; Zero-shot learning; attention layer; Grid search; Adaptive learning rate.*

*** Mohammad Ali Zare Chahooki: chahooki@yazd.ac.ir**

۱. مقدمه

پیشرفت یادگیری عمیق سبب شد کاربرد شبکه‌های عصبی مانند شبکه‌های عصبی بازگشتی[1] و پیچشی در وظایف مختلف پردازش زبان طبیعی مورد توجه قرار گیرد. با این حال، عملکرد مدل‌های یادگیری عمیق در پردازش زبان طبیعی یا در سایر زمینه‌های داده‌کاوی بسیار وابسته به وجود دادگان غنی برچسب‌دار است. از سوی دیگر، در بسیاری از دامنه‌های مرتبط با پردازش زبان طبیعی مانند تحلیل احساسات رمزارزها یا بررسی واکنش مخرب داروها دادگان غنی برچسب‌دار برای آموزش شبکه‌های عصبی عمیق در دسترس نمی‌باشد. بنابراین، آموزش مدل‌های عمیق با مجموع دادگان کوچک سبب افت کارایی این مدل‌ها می‌شود [۱].

یکی از پیشرفت‌های موثر در حوزه پردازش زبان طبیعی و بینایی ماشین یادگیری انتقالی می‌باشد. یادگیری انتقالی در بالابردن کارایی وظایف بینایی ماشین بسیار مؤثر بوده است. دلیل این موفقیت در دسترس‌بودن مجموعه داده‌های بزرگ برچسب‌گذاری شده مانند ایمیج‌نت و کوکو می‌باشد. در نتیجه شبکه‌های عصبی کانولوشن محور بر روی این مجموع دادها به عنوان ستون فقرات شبکه (لایه استخراج‌گر ویژگی) آموزش می‌بینند . یادگیری انتقالی در پردازش زبان طبیعی تا سال ۲۰۱۸ که ترانسفورمرها توسط گوگل معرفی شدند، چندان مورد توجه قرار نگرفته بود [۲]. یکی از وظایف پردازش زبان طبیعی که استفاده از یادگیری انتقالی در آن بسیار مورد توجه قرار گرفته است، تحلیل احساسات می‌باشد. با ظهور شبکه‌های اجتماعی بلادرنگ مانند توییتر، مجموعه‌های بی‌نظیری از دیدگاه‌های عمومی و تخصصی در دسترس قرار گرفته است [۳]. با این حال فقدان داده‌های برچسب‌دار در حوزه‌های خاص مانند رمزارزها سبب شده تحلیل احساسات در این دامنه‌ها چالش برانگیز باشد.

همچنین، مشکل مذکور در وظیفه تحلیل احساسات جنبه‌محور، تشدید می‌شود. زیرا، دادگان بسیار کمیابی با محتوی سهام و یا رمزارزها برای آموزش مدل‌های عمیق حاوی برچسب جنبه‌ها و احساسات هر جنبه وجود دارد. یادگیری انتقالی می‌تواند از دانش موجود در سایر دامنه‌ها برای انجام وظایف خاص در دامنه‌ای مشابه استفاده کند. با این حال استفاده از دادگان در دامنه‌های مختلف که به عنوان دادگان میان دامنه‌ای شناخته می‌شود، نیز چالش بر انگیز است. نشان داده شده استفاده از دادگان نامشابه یا با مشابهت پایین با وظیفه یادگیری دامنه هدف، سبب بروز پدیده انتقال منفی[2] می‌شود. انتقالی منفی به دانش منتقل شده از دامنه مبدا گفته می‌شود که سبب افت کارایی مدل در دامنه هدف شود [۲].

برای انجام وظایف پردازش زبان طبیعی درابتدا تکنیک‌های مانند ایجاد کیسه‌ای از کلمات به وجود آمد که از N-gram استفاده می‌کردند. با این حال، در مدل‌های مبتنی بر کیسه کلمات، معنای کلی متن نادیده گرفته می‌شود. از این رو، روش‌های مبتنی بر تعبیه کلمات ارائه گردید. این روش‌ها، کلمات را از نظرتشابه معنای، دارای زمینه‌های مشابه در نظر می‌گیرند. نقطه ضعف این روش‌ها نیازمندی‌های آنها به مجموعه واژگان گسترده می‌باشد [۴]. معرفی ترانسفورمرها از اوایل سال ۲۰۱۸ کمک شایانی به حل مسائل پردازش زبان طبیعی کرده است. یکی از معروف‌ترین ترانسفورمرها، نمایش‌های رمزگذار دوطرفه از ترانسفورمرها (برت) می‌باشد که بر روی مجموعه بزرگی از داده‌های انگلیسی آموزش دیده است. برت دارای لایه‌های رمزگذاری می‌باشد که با مکانیزم خود توجهی تقویت شده‌اند [۳].

[1] Recurrent Neural Networks

[2] Negative transfer

با این حال، همچنان یکی از چالش‌های اصلی عدم وجود دادگان در دامنه‌های خاص برای آموزش مدل‌های عمیق پایه و پیش‌آموزش دیده می‌باشد. بلو و همکاران [۴] برای افزایش غنای دادگان هدف ستون‌های مربوط به متن‌ها از دادگان مختلف را با یکدیگر الحاق کرده‌اند. در این مقاله نویسندگان، توجهی به مشابهت دادگان هدف و مبدا نکرده و صرفاً با الحاق داده‌ها سعی در افزایش غنای دادگان مبدا داشته‌اند. در این پژوهش به دنبال ارائه پاسخی به سوالات پژوهشی زیر می‌باشیم:

۱- چگونه می‌توان دادگان میان دامنه‌ای مناسبی در دامنه مبدأ انتخاب کرد که انتقال ویژگی‌های آنها به دامنه هدف موجب بروز پدیده انتقال منفی نگردد؟

۲- چگونه می‌توان با ترکیب شبکه‌های یادگیری عمیق، لایه توجه و آموزش آنها با دادگان میان‌دامنه‌ای مدلی مطلوب با کمترین افت کارایی برای وظیفه تحلیل احساسات جنبه‌محور ارائه داد؟

۳- با توجه به اینکه استفاده از مدل‌های یادگیری عمیق پایه و پیش‌آموزش‌دیده برای استخراج ویژگی و انتقال دانش از دامنه مبدأ به هدف مؤثر می‌باشد. کدامیک از این مدل‌ها برای تحلیل احساسات رمزارزها اثربخشی بیشتری دارد؟

۴- دادگان نامتوازن سبب جهت‌گیری مدل به سمت قطبیتی خاص می‌شود. چگونه می‌توان با این مشکل برخورد کرد؟

برای پاسخ به این سوالات در این پژوهش از مدل‌های مختلف یادگیری عمیق پایه، پیش‌آموزش‌دیده و ترکیب آنها با لایه توجه استفاده شده است. بخشی از این پژوهش را می‌توان مقایسه عملکرد مدل‌های متداول یادگیری عمیق (پایه، پیش آموزش‌دیده و ترکیبی) در تحلیل احساسات توئیت‌های مرتبط با رمزارزها دانست. با این حال صرفاً این پژوهش مقایسه مدل‌های پیشین نمی‌باشد، زیرا از روش‌های تنظیم دقیق برای پیاده‌سازی معماری ستون فقرات شبکه‌های عصبی عمیق ترکیبی استفاده شده است. در اینجا، دو روش تنظیم دقیق برای مدل‌های پیش‌آموزش‌دیده و شبکه‌های عصبی عمیق ترکیبی استفاده شده است. روش اول آموزش کل معماری عمیق می‌باشد که شامل بروزرسانی وزن‌های ستون فقرات شبکه و شبکه عصبی عمیق پایانی می‌باشد. روش دوم انجماد کل وزن‌های ستون فقرات شبکه و آموزش شبکه عصبی عمیق پایانی است. مقایسه روش‌های مختلف تنظیم دقیق شبکه‌های عصبی عمیق تاکنون کمتر مورد توجه قرار گرفته است [۵]. لازم به توضیح است که در اینجا ستون فقرات شبکه که همان لایه استخراج‌گر ویژگی‌ها می‌باشد، شبکه‌ای پیش‌آموزش‌دیده از خانواده برت می‌باشد.

همچنین در این پژوهش تاثیر دادگان میان‌دامنه‌ای بر کارایی شبکه‌های عصبی عمیق ترکیبی در یادگیری انتقالی مورد توجه قرار گرفته است. لذا، در این پژوهش برای غنای دادگان رمزارزها در حوزه هدف از دادگانی میان‌دامنه‌ای مشابه در دامنه مبدا استفاده شده است. در این مقاله از یادگیری انتقالی شات صفر[3]، استفاده شده، زیرا مدل در هنگام آموزش هیچ نمونه داده‌ای از دادگان دامنه هدف را نمی‌بیند. در این مقاله بر خلاف مقالات مشابه [۴] که تنها دادگان مختلف را با یکدیگر الحاق می‌کنند، ابتدا شباهت دادگان با توئیت‌های رمزارزها با استفاده از معیار شباهت پیرسون محاسبه شده و سپس دادگان با شباهت بیشتر در دامنه مبدا مورد استفاده قرار می‌گیرد [۶]. در این پژوهش شباهت سه مجموعه دادگان تحلیل احساسات توییتر[4]، احساسات توئیت‌های بیت‌کوین[5] و IMDB با دیتاست دامنه هدف که توئیت‌های جمع‌آوری رمزارزها می‌باشد مقایسه شده

[3] Zero-shot transfer learning
[4] Twitter Sentiment Analysis (kaggle.com)
[5] https://data.world/mercal/btc-tweets-sentiment

است. سپس متون دادگانی که دارای شباهت بیشتر از ۴۰ درصد می‌باشد، با یکدیگر الحاق می‌شوند. همچنین، یکی از مشکلات دادگان در دنیای واقعی عدم توازن قطبیت‌ها می‌باشد. برای مقابله با این مشکل، در این پژوهش از تکنیک روش نمونه‌برداری بیش از حد اقلیت مصنوعی (SMOTE)[6] [۷] استفاده شده است. SMOTE تکنیکی آماری است که به طور مصنوعی نمونه‌های جدیدی از کلاس اقلیت را در فضای ویژگی ایجاد می‌کند تا مجموعه داده را متعادل کند.

به صورت کلی در این مقاله، روشی ترکیبی مبنی بر یادگیری عمیق انتقالی شات صفر برای تحلیل احساسات جنبه‌محور رمزارزها معرفی شده است. در این روش، ابتدا توئیت‌های افراد تاثیرگذار در حوزه رمزارز استخراج شده و عملیات پیش‌پردازش مانند یکسان‌سازی، حذف کلمات توقف، بن‌سازی و حذف کارکترهای زائد روی آنها اعمال می‌شود. سپس دادگان میان‌دامنه‌ای مشابه توئیت‌ها که دارای ضریب همبستگی پیرسون بیشتر از ۴۰ درصد است ادغام شده و پس از عملیات پیش‌پردازش و استفاده از تکنیک SMOTE برای متوازن‌سازی داده‌ها به لایه استخراج‌گر ویژگی (ستون فقرات) یا همان دامنه مبدا برای استخراج دانش تزریق می‌شود. این لایه از شبکه پیش‌آموزش دیده دیستیل‌برت به همراه لایه توجه تشکیل شده است. لایه توجه به مدل کمک می‌کند ویژگی‌های مهم متن را شناسایی کرده و ویژگی‌های غیرضروری را نادیده بگیرد [۸]. پس از استخراج دانش از دادگان میان‌دامنه‌ای این ویژگی‌ها به لایه پایانی یا همان دامنه هدف تزریق می‌شود. در دامنه هدف برای استخراج جنبه از دادگان توئیت‌ها از روش spaCy استفاده شده است [۹]. همچنین، در این لایه برای تحلیل احساسات جنبه‌محور از شبکه عصبی عمیق حافظه کوتاه مدت بلند دو طرفه (BiLSTM) با ترکیب لایه توجه بهره‌گیری شده است. در ادامه، بعد از تزریق دادگان هدف (توئیت‌های اثرگذاران) به مدل ترکیبی تنظیم دقیق شده BiLSTM و لایه توجه، ویژگی‌های اثرگذار بر قطبیت هر جنبه استخراج شده و بعد از عملیات برون‌ریزی[7] این ویژگی‌ها به لایه‌ای کاملاً متصل تزریق می‌شود. در این لایه، با استفاده از تابع سافت‌مکس قطبیت هر جنبه تعیین‌شده و در خروجی نمایش داده می‌شود. در زیر نوآوری‌های اصلی این پژوهش تشریح شده است.

۱- در این پژوهش بر خلاف اغلب مقالات [۱۰-۱۲] که از تحلیل احساسات سنتی با روش‌های عمومی مانند ویدر[8] برای تحلیل احساسات رمزارزها بدون در نظر گرفتن جنبه‌های موجود در متن استفاده کرده‌اند، از یادگیری عمیق انتقالی با مدل‌های تنظیم دقیق شده برای تحلیل احساسات رمزارزها استفاده شده است. همچنین در این مطالعه، از یادگیری انتقالی شات صفر استفاده شده، زیرا مدل هیچ نمونه‌ای در دادگان دامنه هدف را در فرایند آموزش نمی‌بیند.

۲- مدل‌های ترکیبی مورد استفاده توسط سه مجموعه دادگان با شباهت بیشتر از ۴۰ درصد با دادگان رمزارزها آموزش دیده‌اند. برای سنجش شباهت دادگان مبدا و هدف از ضریب همبستگی پیرسون به دلیل اینکه توزیع آماری دادگان کاملاً نرمال نیست، استفاده شده است. این‌کار به صورت موثری انتقال منفی را کاهش می‌دهد. همچنین، از تکنیک SMOTE برای برخورد با پدیده عدم توازن دادگان استفاده شده است.

۳- در این پژوهش دو روش مختلف تنظیم دقیق مدل‌های پیش‌آموزش‌دیده برای شناخت بهترین رویکرد جهت

[6] Synthetic Minority Over-sampling Technique

[7] Dropout
[8] Vader

تحلیل احساسات جنبه‌محور رمزارزها مورد مقایسه قرار گرفته است. همچنین، در این مطالعه از لایه الحاق به عنوان یک سوئیچ استفاده شده است. این لایه کمک می‌کند که بتوان بدون پیاده‌سازی مجدد شبکه‌های عصبی عمیق پایه و پیش‌آموزش دیده را با یکدیگر مقایسه کرد. در واقع با این لایه می‌توان به سرعت قسمت ستوان فقرات شبکه را با مدلی دیگر جایگزین کرد.

۴- در این پژوهش، انتخاب برخی از ابرپارامترها و نرخ یادگیری شبکه‌های عصبی عمیق به صورت وفق‌پذیر و با کمک تکنیک جستجوی گریدی انجام شده است.

۵- در این پژوهش با استفاده از Api توئیتر نظرات تاثیرگذارن در حوزه رمزارزهای مختلف در بازه زمانی هشت‌ماه از سال ۲۰۲۱ تا ۲۰۲۲ جمع‌آوری شده است. این توئیت‌ها شامل تحلیل‌ها و نظرات تخصصی صاحبنظران در حوزه رمزارز می‌باشد و غنای محتوایی آن بسیار بیشتر از استخراج نظرات به وسیله هشتگ‌ها بدون در نظر گرفتن تخصص نویسنده است.

در بخش دوم به در ابتدا مفاهیم شبکه‌های عصبی عمیق پایه، پیش‌آموزش‌دیده و یادگیری عمیق انتقالی را تشریح می‌کنیم؛ سپس، به مرور مقالات پیشین در زمینه تحلیل احساسات جنبه‌محور با یادگیری عمیق انتقالی می‌پردازم. در بخش سوم پیاده‌سازی شبکه‌های عصبی عمیق پیش‌آموزش‌دیده و ترکیبی را شرح می‌دهیم و در انتها نتایج آزمایشات تجربی، نتیجه‌گیری و پیشنهاداتی جهت تحقیقات آینده را بیان می‌کنیم.

۲. مفاهیم و مروری بر تحقیقات پیشین

در این بخش در ابتدا مروری کوتاه بر مفاهیم تحلیل احساسات جنبه‌محور، شبکه‌های عصبی عمیق، پیش‌آموزش‌دیده و یادگیری عمیق انتقالی خواهیم داشت و سپس تحقیقات پیشین در زمینه تحلیل احساسات جنبه‌محور را بررسی می‌کنیم.

۱،۲. مفاهیم تحلیل احساسات جنبه‌محور، شبکه‌های عصبی عمیق و یادگیری عمیق انتقالی

با گسترش شبکه‌های اجتماعی تقریبا بیش از نیمی از جمعیت دنیا در یک یا چند شبکه اجتماعی حضور دارند [۵]. علاوه بر این، رسانه‌های اجتماعی بستری تعاملی جهت تبلیغات، ارائه نظرات تخصصی و کسب وکار فراهم آورده‌اند. این امر سبب شده بسیاری از افراد در تصمیم‌گیری جهت انتخاب کالا یا فرصت سرمایه‌گذاری به شدت به محتوای این رسانه‌ها متکی باشند [۱۳]. توئیتر یکی از شبکه‌های اجتماعی است که به دلیل دسترسی آسان به توئیت‌ها به وسیله API و طول حداکثری ۲۸۰ کاراکتر برای هر توئیت، یکی از کابردی‌ترین میکروبلاگ‌ها برای تحلیل و بررسی نظرات می‌باشد. سینگ و همکاران [۱۴] نشان دادند نظرات کاربران در اغلب اوقات شامل "هدف" و "احساس" است. به عنوان مثال، "قیمت سهام A (هدف) در حال افزایش است (احساس)"، بنابراین تحلیل احساسات شامل شناسایی هدف و احساسات مرتبط با هدف می‌باشد. هدف می‌تواند سند، جمله (سطح جمله) یا حتی جنبه‌ای در یک جمله باشد.

تحلیل احساسات را می‌توان از نظر مدل‌های یادگیری ماشین برای حل مسئله به سه رویکرد دسته طبقه‌بندی کرد: (۱) مبتنی بر واژگان، که از فرهنگ لغتی از پیش تهیه شده برای تشخیص احساسات در سطح کلمه استفاده می‌کند. (۲) یادگیری ماشین پایه، که از الگوریتم‌های طبقه‌بندی کننده‌ای که شبکه عصبی نیستند برای طبقه‌بندی کلمات به برچسب‌های از پیش تعیین‌شده استفاده می‌کند. (۳) یادگیری عمیق، که از شبکه‌های عصبی عمیق(پایه یا پیش‌آموزش‌دیده) برای استخراج ویژگی‌های معنایی و تحلیل احساسات استفاده می‌کند [۱۵]. از سوی دیگر،

تحلیل احساسات از نظر عمق بررسی متن به چهار حالت تحلیل احساسات در سطح سند، جمله، عبارت و جنبه تقسیم می‌شود. در سطح سند، کل متن به‌عنوان موجودیتی واحد در نظر گرفته می‌شود و احساسات عمومی آن بررسی می‌شود؛ در سطح جمله، احساسات موجود در هر جمله به‌طور مجزا تحلیل می‌شود؛ در سطح عبارت، هر عبارت یا تکه‌ای از جمله مورد بررسی قرار می‌گیرد؛ و در سطح جنبه، جنبه‌های متن در سطح هر جمله یا عبارت شناسایی شده و قطبیت هر جمله تعیین می‌شود [۱۶]. در این پژوهش هدف ما تحلیل احساسات جنبه‌محور می‌باشد. تحلیل احساسات جنبه‌محور از نظر نوع وظایف به دو دسته تک و چند وظیفه‌ای تقسیم می‌شود. لیو و همکاران[9] [۱۷] بیان می‌کنند جمله دارای چهار المان احساسی برای استخراج می‌باشد که عبارتند از : کلمه جنبه[10]، طبقه یا گروه جنبه[11]، کلمه عقیده[12] و قطبیت احساسات[13]. برای مثال در جمله "بیت‌کوین فوق‌العاده است"، کلمه جنبه برابر بیت‌کوین، گروه جنبه برابر رمزارز، کلمه عقیده برابر فوق‌العاده و قطبیت احساسات برابر مثبت می‌باشد. با این چهار المان در تحلیل احساسات جنبه‌محور تک وظیفه‌ای، می‌توان چهار وظیفه تعریف کرد، به عبارت دیگر استخراج هر المان در جمله یک وظیفه محسوب می‌شود. همچنین، در تحلیل احساسات جنبه‌محور چند وظیفه‌ای، می‌توان ترکیب‌های دوتایی، سه‌تایی و یا استخراج کل المان‌ها را به عنوان (چهارتایی) یک وظیفه تعریف کرد، که در این پژوهش ترکیب‌های دوتایی استخراج جنبه و قطبیت آن مورد مطالعه قرار گرفته است.

۱،۲،۲. شبکه‌های عصبی عمیق

یادگیری عمیق در واقع، زیر مجموعه‌ای از یادگیری ماشین می‌باشد. یادگیری عمیق نیازی به قوانین طراحی شده توسط انسان ندارد، بلکه از مقدار زیادی داده برای نگاشت ورودی به برچسب‌های از پیش تعریف‌شده (در مسائل طبقه‌بندی) استفاده می‌کند [۱۸]. مدل‌های پایه یادگیری ماشین نیازمند چندین مرحله متوالی پیش‌پردازش، استخراج و انتخاب ویژگی‌های برتر، یادگیری و طبقه‌بندی می‌باشند. در حالیکه، مدل‌های یادگیری عمیق، توانایی یادگیری مجموعه‌ای از ویژگی‌ها به صورت خودکار را دارا می‌باشند [۱۹].

یکی از انواع خانواده مدل‌های عصبی که بسیار در حل مسائل پردازش زبان طبیعی مورد استفاده قرار گرفته، شبکه‌های عصبی بازگشتی (RNNs) می‌باشند. ویژگی برجسته این شبکه‌ها استفاده از "حافظه" می‌باشد. حافظه این امکان را به شبکه می‌دهد که اطلاعات ورودی‌های قبلی را برای اثرگذاری در ورودی فعلی استفاده کند [۲۰]. به بیان رسمی اگر "x" لایه ورودی، "h" لایه پنهان و "y" لایه خروجی شبکه باشد. در هر زمان t، ورودی فعلی ترکیبی از ورودی $x(t)$ و $x(t-1)$ حاصل می‌شود. بنابراین، خروجی در هر زمان t برای بهبود خروجی بعدی ($t+1$) به شبکه برمی‌گردد. با این حال این شبکه‌ها دارای دو مشکل اساسی محوشدگی و انفجار گرادیان می‌باشند. گرادیان‌ها اطلاعات مورد استفاده در شبکه عصبی بازگشتی را حمل می‌کنند. زمانیکه گرادیان خیلی کوچک شود، بروزرسانی پارامترها غیر موثر می‌شود و به این پدیده محوشدگی گرادیان می‌گویند. از سوی دیگر، در حین آموزش شبکه عصبی، اگر شیب به جای کاهش، تمایل به رشد تصاعدی داشته باشد، پدیده انفجار گرادیان رخ می‌دهد. این مشکل زمانی ایجاد می‌شود که خطای بزرگ درگرادیان‌ها انباشته شده و سبب می‌گردد بروزرسانی بسیار بزرگی در وزن‌های شبکه عصبی در حین فرآیند آموزش رخ دهد [۲۰].

برای رفع مشکل RNN، شبکه‌های عصبی عمیق حافظه بلند کوتاه مدت (LSTM)، واحد بازگشتی گیتی (GRU) و

[9] Liu et al.
[10] Aspect Term
[11] Aspect Category
[12] Opinion Term
[13] Sentiment Polarity

BiLSTM، معرفی شده‌اند. در واقع می‌توان LSTM را مدل پایه دو شبکه دیگر نامید که با اعمال تغییراتی در LSTM به وجود آمده‌اند. LSTM نوع خاصی از شبکه عصبی بازگشتی بوده و قادر به یادگیری وابستگی‌های طولانی مدت می‌باشد. LSTM از سه قسمت تشکیل شده و هر قسمت عملکرد جداگانه‌ای دارد. بخش اول انتخاب می‌کند که آیا اطلاعاتی که از دوره زمانی *t-1* به دست آمده باید به خاطر سپرده شود یا فراموش گردد. در قسمت دوم، LSTM سعی می‌کند اطلاعات ورودی جدید را بیاموزد و در نهایت در بخش سوم، LSTM اطلاعات بروز شده دوره زمانی فعلی (t) را به دوره زمانی بعدی ($t+1$) منتقل می‌کند. این سه قسمت در LSTM به ترتیب به عنوان دروازه‌های فراموشی، ورودی و خروجی شناخته می‌شوند [۲۰]. نوعی دیگری از شبکه عصبی بازگشتی که تا حد زیادی مشابه LSTM می‌باشد و برای حل مشکل حافظه کوتاه مدت پیاده‌سازی شده، GRU نام دارد. در این شبکه جهت تسریع در فرایند آموزش و سبکسازی مدل از دو دروازه بجای سه دروازه استفاده شده است. این دروازه‌ها که به نام‌های تنظیم مجدد و به‌روزرسانی شناخته می‌شوند. این دروازه‌ها، همانند دروازه‌های LSTM، کنترل می‌کنند چه میزان و کدام اطلاعات باید حفظ شود [۲۶]. نوعی دیگری از LSTM که اجازه پردازش اطلاعات در دوجهت را می‌دهد BiLSTM نام دارد. برخلاف LSTM، در این شبکه ورودی در هر دو جهت (رو به عقب و رو به جلو) جریان دارد. به عبارت دیگر، BiLSTM یک لایه LSTM اضافه دارد که جهت جریان اطلاعات را معکوس می‌کند. به این معنی‌که، دنباله ورودی در لایه اضافی LSTM به عقب جریان می‌یابد؛ سپس، خروجی‌های هر دو لایه LSTM با روش‌های مانند میانگین، مجموع، ضرب یا الحاق ترکیب می‌شوند [۲۰].

۲،۲،۲. یادگیری انتقالی عمیق و شبکه پیش‌آموزش‌دیده برت

یادگیری انتقالی تکنیکی است که از دانش موجود در دامنه مبدا برای بهبود عملکرد وظیفه در دامنه هدف استفاده می‌کند. در یادگیری انتقالی دامنه مبدا را با D و دامنه هدف را با T نشان می‌دهند. دامنه مبدا (D) از دو جزء تشکیل شده است: داده که با X و توزیع آن که با P(X) و به صورت رابطه (۱) نشان می‌دهیم [۲۱].

$$D = X, P(X) \quad (۱)$$

دامنه هدف (T) نیز از دو جزء تشکیل شده است: داده برچسب y و تابع پیش‌بینی $f(.)$ که به صورت رابطه (۲) نشان می‌دهیم.

$$T = y, f(.) \quad (۲)$$

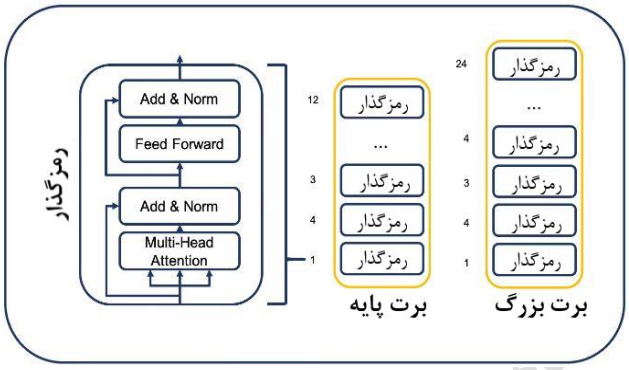
به بیان رسمی، یادگیری انتقالی رویکردی است برای بهبود وظیفه یادگیری (TT) در دامنه هدف (TD)، با دانش آموخته‌شده از وظیفه یادگیری (ST) در دامنه مبدا (SD)[۱۴]، که معمولاً $SD \neq TD$ و $ST \neq TT$ می‌باشد. شبکه‌های عصبی پیش‌آموزش‌دیده مدل‌های مناسبی جهت پیاده‌سازی تکنیک‌های یادگیری انتقالی می‌باشند [۲۱]. یادگیری عمیق انتقالی را می‌توان به سه دسته زیرو، تک و چند شات[۱۵] تقسیم کرد. در یادگیری عمیق انتقالی زیروشات که در این مقاله نیز استفاده شده است، مدل هیچ نمونه‌ای از داده‌های هدف دارای برچسب را در اختیار ندارد و تنها با دادگان میان دامنه‌ای مشابه آموزش دیده است. در یادگیری انتقالی تک‌شات، مدل تنها یک نمونه از داده‌های هدف دارای برچسب را در اختیار دارد. از سوی دیگر، در یادگیری انتقالی با شات محدود مدل تعداد کمی از نمونه‌های داده‌های هدف دارای برچسب را در اختیار دارد و باید بتواند با استفاده از این تعداد کم نمونه و دانش قبلی خود، پیش‌بینی‌های دقیقی

[14] TT: Target Task, TD: Traget Domain, ST: Source Task, SD: Source Domain

[15] Zero/One/Few Shot transfer Learning

انجام دهد. برای انجام یادگیری انتقالی، در سال‌های اخیر شبکه‌های عصبی عمیق پیش‌آموزش‌دیده خانواده برت در پردازش زبان طبیعی استفاده شده است. برت مبنی بر یادگیری انتقالی بوده از ترانسفورمرها استفاده می‌کند. یادگیری انتقالی مستلزم آموزش مدل برای وظیفه‌ای با مجموعه دادگان غنی در دامنه مبدا و انتقال دانش به دست آمده برای وظیفه دیگری در دامنه هدف می‌باشد [۲۲].

برت، برای دو وظیفه زبان پوشانده‌شده و بدون پوشش[16] آموزش دیده است. در وظیفه زبان پوشانده‌شده، جملات به شبکه داده می‌شود و برخی از کلمات برای مدل پنهان می‌شوند. سپس، مدل تلاش می‌کند این کلمات پنهان را پیش‌بینی کند. در وظیفه زبان بدون پوشش، دو جمله به مدل داده می‌شود و مدل باید پیش‌بینی کند یک جمله می‌تواند ادامه جمله دیگری باشد یا خیر. مدل برت به صورت پیش‌فرض بر روی دو مجموعه داده بزرگ ویکی پدیای انگلیسی و مجموعه دادگان کتاب‌ها که شامل ۱۱،۰۳۸ کتاب می‌باشد، آموزش دیده است. این مدل، از رمزگذاری مبتنی بر ترانسفورمرها که نوعی شبکه عصبی می‌باشند استفاده می‌کند. برت یک جمله را به عنوان ورودی می‌گیرد و تمامی کلمات جمله را نشانه‌گذاری[17] می‌کند. سپس، کلمات نشانه‌گذاری‌شده به مدل برت باز می‌گردد. خروجی برت نمایش برداری برای هر کلمه نشانه‌گذاری شده است [۲۲]. مدل‌های که از لایه‌های رمزگذارهای مانند برت استفاده می‌کنند، می‌توانند درک بهتری از زمینه متن داشته باشد. زیرا رمزگذار، کل جمله را به عنوان ورودی و به صورت همزمان پردازش می‌کند. این کار سبب می‌شود هنگام تعیین زمینه‌ای برای یک کلمه، برت ورودی‌های قبلی و بعدی کلمه را نیز در نظر بگیرد. در حالیکه شبکه‌های مبتنی بر بازگشت مانند LSTM، ورودی فعلی را فقط با در نظر گرفتن ورودی‌های قبلی پردازش می‌کنند. برت به صورت استاندارد به دو شکل پایه و بزرگ ارائه شده است. برت پایه شامل ۱۲ لایه رمزگذار و ۷۶۸ لایه پنهان است. از سوی دیگر، برت بزرگ دارای ۲۴ لایه رمزگذار با ۱۰۲۴ لایه پنهان می‌باشد. انواع دیگر مدل‌های برت که اغلب مدل‌های فشرده‌تر، سریعتر و موثرتر از برت پایه می‌باشند، عبارتند از: آلبرت، الکترا، دستیل‌برت، اسپن‌برت، برت‌سام[18] و XLNet. در این پژوهش از سه مدل پایه برت، XLNet و دستیل‌برت استفاده شده است. در شکل (۱) تفاوت برت پایه و بزرگ نشان داده شده است.



شکل ۱- مقایسه دو نسخه برت پایه و بزرگ [۲۲]

در ادامه دو نوع شبکه عصبی عمیق مبتنی بر برت یعنی برت تقطیره‌شده و XLNeT بررسی می‌گردد.

دستیل‌برت نسخه سبکسازی شده برت می‌باشد. این مدل فشرده‌تر، کوچک‌تر، سریعتر و سبک‌تر از برت است. استفاده از برت به دلیل وجود میلیون‌ها پارامتر برای کاربردهای دنیای واقعی به خصوص در تلفن‌های همراه یا دستگاه‌ها با منابع پردازشی ضعیف، چالش برانگیز است. اگر چه ممکن است معمارهای پیچیده‌تر به دقت بالاتری دست یابند، اما می‌شود با روش‌های تنظیم دقیق به میزان قابل توجهی دقت مدل‌های سبک‌تر را نیز بهبود داد. در مقایسه با مدل‌های مانند برت بزرگ و XLNet، که عملکرد مناسبی دارند، هدف دستیل‌برت کاهش زمان محاسبه است. در فشرده‌سازی دستیل‌برت، از چارچوب

[16] Masked and Unmasked Language
[17] Tokenization
[18] AlBERT, SpanBERT, ELECTRA, SumBERT

«معلم-دانش‌آموز» استفاده شده که به آن تقطیر دانش نیز گفته می‌شود. در این تکنیک، مدل یا شبکه‌ای بزرگ به عنوان «معلم» آموزش داده می‌شود و دانش به مدل کوچک‌تر که «دانش‌آموز» نامیده می‌شود منتقل می‌گردد [۲۳]. در دستیل‌برت، توکن‌های تعبیه شده حذف شده و همچنین تعداد پارامترهای شبکه کاهش تعداد یافته است. این کار سبب شده تا حد زیادی محاسبات شبکه کاهش یافته و سرعت یادگیری بهبود یابد. بر اساس استاندارد ارزیابی عمومی مدل‌های زبانی (GLUE)، کارایی دستیل‌برت بدون هیچ‌گونه تنظیم دقیق ۹۷ درصد برت پایه می‌باشد. با این حال، ۴۰ درصد پارامترهای شبکه عصبی کاهش‌یافته و زمان آموزش به میزان قابل‌توجهی سریعتر شده است. محققان نشان داده‌اند [۲۳] با ترکیب روش تقطیر انتقالی و چارچوب یادگیری دو مرحله‌ای، می‌توان به دقتی برابر و یا بالاتر از برت پایه دست یابند. در ادامه XLNeT را معرفی می‌کنیم.

XLNet در واقع نوعی برت است که با ترکیب ترانسفورمرهای بزرگ دو جهته -XL و رمزگذاری خودکار برت، ایجاد شده است. در این شبکه از تکنیک جایگشت برای یافتن زمینه متن به صورت دو جهته استفاده شده، همچین نشانه‌ها به طور تصادفی پیش‌بینی می‌شوند. در برت پایه ۱۵٪ از توکن‌ها پوشانده شده و سایر نشانه‌ها به جای ترتیب تصادفی به صورت متوالی پیش‌بینی می‌شوند. لذا، پیش‌بینی تصادفی نشانه‌ها به XLNet امکان می‌دهد وابستگی‌ها و روابط طولانی بین کلمات را بهتر یاد بگیرد. تحقیقات نشان داده XLNet پایه در وظایف پردازش زبان طبیعی از جمله تحلیل احساسات، پاسخ به سؤال و استنتاج زبان طبیعی بهتر از برت پایه عمل می‌کند. به عبارت دیگر، تکنیک زمینه دو جهته به کلمات قبل و بعد از نشانه نگاه می‌کند تا پیش‌بینی دقیق‌تری ارائه دهد [۲۴].

۲،۲. پیشینه تحقیق در تحلیل احساسات شبکه‌های اجتماعی با توجه به رمزارزها

همانگونه که ذکر شد رویکردهای حل مسئله در تحلیل احساسات از نظر مدل‌های پیاده‌سازی می‌توان به سه دسته مبتنی بر واژگان، یادگیری ماشین پایه و یادگیری عمیق با شبکه‌های عصبی عمیق(پایه، ترکیبی یا پیش‌آموزش دیده) تقسیم کرد. در ادامه پژوهش‌های مرتبط با این تقسیم‌بندی‌های یادشده را بررسی می‌کنیم.

۱،۲،۲. تحلیل احساسات با یادگیری ماشین پایه

سه‌حسینی و همکاران (۲۰۲۰) [۲۵] از بیز ساده و K-همسایه نزدیک برای تعیین قطبیت احساسات توئیت‌ها استفاده کرده‌اند. در این مطالعه نشان داده شد، بیز ساده کارایی بالاتری نسبت به از K-همسایه نزدیک دارد. در پژوهشی دیگر (۲۰۲۱) [۳] توئیت‌های مرتبط با بررسی محصولات آمازون جمع‌آوری شده و سپس سه مدل یادگیری ماشین‌های بردار پشتیبان، لجستیک رگرسیون و K-همسایه نزدیک، برای تحلیل احساسات داده‌ها با یکدیگر مقایسه شده‌اند. نتایج آزمایشات نشان می‌دهد مدل لجستیک رگرسیون با دقت بیش از۸۸ درصد بهترین کارایی را داشته است.

نویسندگان در پژوهشی (۲۰۱۸) [۲۶]، روشی برای پیش‌بینی تغییرات قیمت بیت‌کوین و اتریوم با استفاده از داده‌های توئیتر و گوگل ترند ارائه کرده‌اند. با تجزیه و تحلیل توئیت‌ها، پژوهشگران متوجه شده‌اند که حجم توئیت‌ها، به جای احساسات توئیت، می‌تواند پیش‌بینی‌کننده جهت قیمت باشد. در این مقاله از مدلی خطی جهت پیش‌بینی تغییرات قیمت استفاده شده است. در پژوهشی دیگر (۲۰۲۱) [۲۷]، از مدل ترکیبی LSTM و GRU با ورودی‌های ارزش بازار سهام، حجم معاملات، ویژگی‌های قیمتی (OHLCV) و حجم نظرات گوگل

ترند و توئیتر جهت پیش‌بینی مونرو و لایت‌کوین، استفاده شده است.

تحقیقات در حوزه LSTM نشان داده، استفاده از این روش به این دلیل توانایی آن در به خاطر سپردن ویژگی‌های داده‌ها، بسیار مؤثر است. در همین راستا هوانگ و همکاران (۲۰۲۱) [۲۸]، مدلی برای تحلیل احساسات مبتنی بر LSTM برای پیش‌بینی قیمت رمزارزها ارائه کرده‌اند. این مدل با بهره‌گیری از قابلیت LSTM در یادگیری وابستگی‌های بلندمدت در داده‌های متنی، به تحلیل احساسات کاربران در پلتفرمی چینی پرداخته است. آن‌ها با ایجاد یک دیکشنری اختصاصی برای واژگان مرتبط با رمزارزها، دقت تحلیل احساسات را بهبود داده و نتایج نشان داد که این رویکرد توانسته است در مقایسه با مدل‌های سنتی خودرگرسیونی، به طور قابل‌توجهی دقت و فراخوان بالاتری در پیش‌بینی نوسانات قیمت رمزارزها داشته باشد. همچنین نیر و همکاران (۲۰۲۴) [۱۱]، به بررسی مدل‌های تحلیل احساسات برای توییت‌های مرتبط با رمزارزها با استفاده از تکنیک‌های مختلف یادگیری عمیق پرداخته‌اند. آن‌ها از پنج مدل یادگیری عمیق شامل RNN، LSTM، GRU و ترکیب BiLSTM و CNN برای تحلیل احساسات توییت‌های مربوط به بیت‌کوین استفاده کرده‌اند. داده‌های مورد استفاده شامل بیش از ۱٫۵ میلیون توییت بوده که از وب‌سایت Kaggle استخراج شده است. در ارزیابی مدل‌ها از ویدر برای محاسبه نمره احساسات و از فست‌تکست[19] برای استخراج ویژگی‌ها استفاده شد. نتایج نشان داد که مدل LSTM با دقت ۹۵٫۹۵٪ بهترین عملکرد را در پیش‌بینی و طبقه‌بندی احساسات توییت‌ها به دست آورده است.

در مقالاتی که از داده‌های توئیتر جهت پیش‌بینی قیمت رمزارز استفاده شده یکی از خلاهای تحقیقاتی عدم تنظیم دقیق مدل برای تحلیل احساسات رمزارز می‌باشد. در این مقالات اغلب پژوهشگران از روش‌های برپایه واژگان مانند ویدر (۲۰۱۴) [۲۹] برای تشخیص خودکار احساسات استفاده می‌کنند. ویدر مدلی برای تحلیل احساسات مبتنی بر فرهنگ لغت و قواعد از پیش تعریف‌شده است که به‌طور خاص برای تحلیل احساسات در متون کوتاه و غیررسمی، مانند نظرات شبکه‌های اجتماعی و دیدگاه‌های برخط، توسعه یافته است. این مدل با استفاده از لغت‌نامه‌ای غنی از کلمات و عبارات احساسی، به هر کلمه یا عبارت یک نمره احساسی (مثبت، منفی، یا خنثی) اختصاص می‌دهد. یکی از ویژگی‌های برجسته ویدر توانایی آن در درک و تحلیل جزئیات احساسی خاص متون غیررسمی است، مانند استفاده از علائم نگارشی، شدت‌سنج‌ها (مانند "خیلی" یا "بسیار") و حتی اصطلاحات عامیانه است. ویدر به دلیل سادگی، دقت بالا و عدم نیاز به داده‌های آموزشی در بسیاری از کاربردهای پردازش زبان طبیعی در محیط‌های واقعی استفاده می‌شود [۲۹]. این روش به‌ویژه در تحلیل احساسات رسانه‌های اجتماعی محبوبیت زیادی پیدا کرده و نتایج قابل اعتمادی را ارائه داده است [۱۰, ۱۲] (۲۰۲۳, ۲۰۲۰). در سوتیریوس و همکاران (۲۰۲۲) [۳۰]، به بررسی امکان پیش‌بینی قیمت ارزهای دیجیتال با استفاده از تحلیل احساسات در شبکه‌های اجتماعی پرداخته شده است. نویسندگان داده‌های توییتر و قیمت هفت ارز دیجیتال محبوب را جمع‌آوری و پس از پاکسازی، با استفاده از ابزار ویدر به تحلیل احساسات پرداختند. سپس، برای ارزیابی تأثیرگذاری احساسات بر قیمت، از آزمون‌های علیت گرنجر استفاده کردند. نتایج نشان داد که در برخی از رمزارزها مانند اتریوم و پولکادات، احساسات توییتری می‌تواند منجر به پیش‌بینی بسیار دقیق‌تر شود. این پژوهش بر تأثیر قابل‌توجه احساسات شبکه‌های اجتماعی بر نوسانات بازار رمزارزها تأکید دارد. با این حال، بدیهی‌است درصد خطای روش‌های مانند ویدر می‌تواند در تحلیل احساسات دامنه‌های خاص بالا باشد، زیرا این روش‌ها اغلب برای توئیت‌های

[19] FastText

عمومی ساخته شده‌اند. همچنین، مدل‌های پایه یادگیری ماشین، نیاز به دادگان غنی برچسب‌دار جهت آموزش دارند که در اکثر اوقات این دادگان موجود نمی‌باشد. علاوه بر این، استخراج ویژگی در مدل‌های پایه یادگیری ماشین برخلاف شبکه‌های عصبی عمیق به صورت دستی می‌باشد. بزرگ‌ترین مشکل در این مدل‌ها عدم امکان استفاده از یادگیری انتقالی می‌باشد.

۲،۲،۲. تحلیل احساسات رمزارزها با مدل‌های یادگیری عمیق

سیهان و همکاران (۲۰۲۲) [۳۱]، رویکردی برای تحلیل احساسات مرتبط با رمزارزها ارائه کرده‌اند. در این مطالعه، معماری ترکیبی عمیق شامل لایه‌های پیچشی ، مکانیسم بهبود گروهی[20]، لایه‌های دو جهته، مکانیسم توجه و لایه‌های کاملاً ارائه شده است. مکانیسم بهبود گروهی به بهبود ویژگی‌های مفید کمک می‌کند، در حالی که مکانیسم توجه با اختصاص وزن‌های مختلف به ویژگی‌ها، اطلاعات زمینه‌ای را برجسته می‌سازد. نتایج نشان می‌دهد معماری پیشنهادی، عملکرد مطلوبی نسبت به معماری‌های پیشین بدست آورده است.

مودیچ و همکاران (۲۰۲۴) [۳۲]، به بررسی استفاده از تکنیک‌های تعبیه گراف برای تحلیل احساسات در زمینه رمزارزها پرداخته‌اند. نویسندگان با هدف بهبود دقت مدل‌های تحلیل احساسات، از روش تعبیه گراف برای نمایش و استخراج روابط معنایی و نحوی بین کلمات در داده‌های متنی استفاده کردند. در این پژوهش، داده‌های مورد استفاده به صورت گراف نمایش داده شد و با بهره‌گیری از تکنیک‌های تعبیه گراف، مانند DeepWalk و مدل Word2Vec، تبدیل به بردارهای کم‌بعد شدند که ویژگی‌های ساختاری متن را حفظ می‌کنند. سپس این بردارها به عنوان ورودی به یک مدل BiLSTM داده شد می‌شود. نتایج می‌دهد، مدل تعبیه گرافی به طور قابل‌توجهی عملکرد بهتری نسبت به مدل‌های سنتی تعبیه کلمات داشته و دقت بالاتری بدست آورده است. این پژوهش نشان می‌دهد استفاده از تعبیه گرافی می‌تواند منجر به بهبود درک و تحلیل دقیق‌تر احساسات در متون مرتبط با رمزارزها شود. همچنین اسالم و همکاران (۲۰۲۲) [۳۳]، مدلی ترکیبی مبنی بر LSTM و GRU برای تحلیل احساسات توییت‌های مرتبط با رمزارزها ارائه کرده‌اند. آن‌ها با جمع‌آوری ۴۰،۰۰۰ توییت رمزارز، این داده‌ها را پس از پیش‌پردازش با استفاده از مدل ترکیبی پیشنهادی تحلیل کردند. برای استخراج ویژگی‌ها از روش‌های BoW، TF-IDF و Word2Vec استفاده شده است و سپس این ویژگی‌ها به مدل ترکیبی LSTM-GRU داده شده‌اند. نتایج نشان این پژوهش بیان می‌کند که استفاده از مدل‌های ترکیبی می‌تواند به طور قابل توجهی دقت تحلیل احساسات رمزارزها را بهبود دهد. در رویکری دیگر که مبنی بر مدل پیش‌آموزش‌دیده روبرتا است، داوچف و همکاران (۲۰۲۰) [۳۴]، برای پیش‌بینی قیمت بیت‌کوین از مجموعه دادگان توئیت‌های کگل[21] و دادگان قیمتی پیشین استفاده کرده‌اند. در این پژوهش قطبیت توئیت‌های مربوط به بیت‌کوین بعد از پیش‌پردازش، توسط شبکه روبرتا تعیین شده است. سپس، مقادیر حاصل از قطبیت توئیت‌ها با مقادیر قیمت‌های گذشته بیت‌کوین ترکیب شده و برای پیش‌بینی قیمت به رگرسورهای FbProphet و XGBoost تزریق شده است.

ژائو و همکاران (۲۰۲۰) [۳۵]، رویکردی را برای تحلیل احساسات جنبه‌محور پیشنهاد کردند که از شبکه‌های پیچشی مبتنی بر گراف[22] برای دریافت وابستگی‌های احساسات بین جنبه‌های مختلف در یک جمله استفاده می‌کند. نویسندگان استدلال می‌کنند که رویکردهای موجود عمدتاً این وابستگی‌ها را نادیده می‌گیرند، که می‌تواند اطلاعات ارزشمندی را برای تحلیل

[20] Group-wise Enhancement Mechanism

[21] Kaggle

[22] Graph Convolutional Networks (GCNs)

احساسات ارائه دهد. مدل ارائه شده شبکه‌های پیچشی مبنی بر گراف وابستگی احساسات (SDGCN) ابتدا از مکانیسم توجه دو طرفه با رمزگذاری موقعیت برای به دست آوردن نمایش‌های خاص جنبه استفاده می‌کند؛ سپس، از GCN برای مدل‌سازی وابستگی‌های احساسات بین جنبه‌ها بهره می‌گیرد. زونیک و همکاران (۲۰۲۱) [۳۶]، رویکردی برای تحلیل احساسات جنبه‌محور با استفاده از GCN بر روی نمودارهای وابستگی نحوی معرفی کرد. نویسندگان استدلال می‌کنند که ساختار نحوی جملات برای تجزیه و تحلیل دقیق احساسات بسیار مهم است، به ویژه برای حوزه‌های سلامت و رفاه که در آن احساسات اغلب به دلیل ماهیت محتوا به سمت منفی گرایش پیدا می کنند. روش آنها شامل ایجاد نموداری از جملات است که در آن گره‌ها کلمات و یال ها وابستگی‌های نحوی را نشان می دهند. در این مطالعه، به اندازه کافی عدم تعادل طبقاتی در مجموعه دادگان را در نظر نمی‌گیرد. این مشکل، می‌تواند توانایی مدل برای تعمیم در کلاس‌های احساسات مختلف را تغییر دهد. علاوه بر این، این مقاله ممکن است فاقد بحث مفصل در مورد کارایی محاسباتی مدل، که جنبه‌ای حیاتی در هنگام کاربرد برای سناریوهای دنیای واقعی است. علاوه بر این، این مطالعه با تمرکز بر یک حوزه خاص (سلامت و رفاه) محدود شده، در نتیجه سوالاتی در مورد عملکرد آن بر روی مجموعه دادگان با زمینه‌های متفاوت ایجاد می‌شود. لی و همکاران (۲۰۲۱) [۳۷]، برای شناسایی موثر جفت‌های جنبه-نظر مدلی بمنی بر برچسب‌گذاری فاصله محور[23] (QDSL) ارائه کرده‌اند. نویسندگان وظیفه استخراج جفت جنبه – نظر[24] (AOPE) را به وظایف فرعی استخراج کلمه جنبه[25] (ATE) و استخراج نظر با جنبه مشخص (ASOE) تقسیم می‌کنند. این مدل از یک طرح برچسب گذاری مبتنی بر دهانه برای شناسایی جفت های جنبه-نظر استفاده می‌کند و پیشرفت‌های قابل توجهی نسبت به روش‌های پیشین در مجموعه داده‌های معیار نشان می‌دهد. در این تحقیق، علاوه بر در نظر نگرفتن مشکل عدم تعادل کلاس در مجموعه داده و عدم تنظیم دقیق مدل پیشنهادی، مدل پیشنهادی بر روی یک مجموعه داده با زمینه متفاوت، آزمایش نشده است. هبت و همکاران (۲۰۲۲) [۳۸]، مدلی ترکیبی از شبکه‌های عصبی بازگشتی مانند LSTM، BiLSTM و GRU و شبکه‌های عصبی کانولوشنی با لایه‌های تعبیه‌شده مختلف (XLNet،MultiFiT و CamemBERT) ارائه کرده‌اند. نتایج نشان می‌دهد که ترکیب GRU و CNN با XLNet در سه مجموعه داده فرانسوی: نظرات مشتریان فرانسوی آمازون (FACR)، AlloCiné Dataset (AC) و تجزیه و تحلیل احساسات فرانسوی توییتر، به ترتیب با دقت ۹۶,۵، ۹۰,۱ و ۸۹,۶ درصد دست یافته است.

همانگونه که در مقالات مبتنی بر مدل‌های پیش‌آموزش‌دیده [۴, ۳۸] (۲۰۲۳, ۲۰۲۲) نشان داده شده، مدل‌های عصبی عمیق باید نسبت به وظیفه و نوع داده ورودی، تنظیم گردند. نشان داده شده (۲۰۲۱)[۲۱] مدل‌های عصبی عمیقی که در تحلیل احساسات دادگان توئیت‌های عمومی و یا نظرات فیلم‌ها کارایی مناسبی دارند، کارایی خود را بر روی دادگانی مانند توئیت‌های اثرات مخرب واکسن و یا دارو از دست می‌دهند. از این رو، در این پژوهش به دلیل کمبود نظرات مرتبط با پیش‌بینی رمزارزهای مختلف از دادگان میان دامنه‌ای با درجه شباهت کسینوسی بیشتر از ۴۰ درصد استفاده شده است. سپس متن‌های این دادگان با هم الحاق شده و به شبکه‌های عصبی عمیق پایه و پیش‌آموزش‌دیده دامنه مبدا جهت استخراج ویژگی تزریق شده است. در ادامه، دانش موجود در دامنه مبدا برای آموزش شبکه عصبی عمیق دامنه هدف استفاده شده است. معماری مورد استفاده این مقاله با دو روش مختلف که در بخش بعد مفصل تشریح می‌گردد،

[23] Question-Driven Span Labeling (QDSL)
[24] Aspect-Opinion Pair Extraction (AOPE)
[25] Aspect Term Extraction (ATE)

تنظیم دقیق شده است. تنظیم دقیق مدل دامنه هدف و انتخاب دادگان مشابه در دامنه مبدا کمک می‌کند کارایی روش پیشنهادی بهبود یافته و از انتقال منفی به شدت کاسته شود. همچنین، در این مقاله، داده‌ها شامل جملات متنی هستند که به منظور تحلیل احساسات جمع‌آوری شده‌اند. هر جمله به عنوان یک نمونه داده در نظر گرفته می‌شود و دارای یک برچسب است که بیانگر احساسات مرتبط با آن جمله (مثبت، منفی یا خنثی) می‌باشد. بنابراین، ماهیت داده‌ها متنی است و هر نمونه داده یک جمله کامل است، نه کلمات یا بخش‌های جداگانه از یک جمله. همانگونه که ذکر شد در اینجا از تکنیک SMOTE برای حل مشکل دادگان نامتوازن استفاده می‌شود. SMOTE به طور معمول برای رفع مشکل عدم توازن در داده‌های عددی و دسته‌بندی‌ها استفاده می‌شود. این تکنیک با تولید نمونه‌های جدید برای کلاس‌های اقلیت، توازن بین کلاس‌ها را برقرار می‌کند. در زمینه داده‌های متنی، SMOTE به طور مستقیم بر روی داده‌های متنی خام (جملات یا کلمات) اعمال نمی‌شود، بلکه ابتدا نیاز است که داده‌های متنی به یک نمایش عددی (نظیر بردارهای ویژگی) تبدیل شوند. در این مقاله، جملات متنی ابتدا با استفاده از توکنایزر روبرتا به توکن‌های کوچک‌تر تبدیل می‌شوند و سپس به بردارهای تعبیه‌ای[26] که نمایانگر ویژگی‌های معنایی عمیق جملات هستند، تبدیل می‌گردند. این بردارها نمایش عددی جملات را تشکیل می‌دهند و هر بردار نماینده یک جمله است. پس از تبدیل داده‌های متنی به این فضای تعبیه‌ای، SMOTE بر روی این بردارهای عددی اعمال می‌شود تا نمونه‌های جدیدی از کلاس‌های اقلیت تولید شود. این نمونه‌های جدید به عنوان بردارهای تعبیه‌ای جدید به مجموعه داده اضافه می‌شوند (۲۰۲۳) [۷].

کبیر و همکاران (۲۰۲۲) [۳۹]، چارچوبی برای طبقه‌بندی احساسات مبتنی بر جنبه را معرفی کردند. نوآوری اصلی این تحقیق در روش ترکیبی آن نهفته است که عبارت است از ترکیبی از سینتکس و رویکردهای مبتنی بر وابستگی برای استخراج جنبه‌ها از متن انتقادات مرتبط با دادگان هتل. این مطالعه به دلیل تمرکز بر یک حوزه واحد (بررسی هتل)، ابهاماتی در مورد کاربرد و اثربخشی مدل پیشنهادی در دامنه‌های مختلف ایجاد می‌کند.

بو هوانگ و همکاران (۲۰۲۲) [۴۰]، مدل CPA-SA را معرفی کردند که محتوای معنایی یک جمله را در نظر می‌گیرد و وزن کلمات متنی را بر اساس موقعیت مربوط به جنبه تنظیم می‌کند. هدف این روش کاهش تأثیر تفاوت تعداد کلمات در هر دو طرف کلمات جنبه هنگام تعیین قطبیت احساسات است. علاوه بر این، این مدل تأثیر ارتباطات متنی بین جملات در یک سند را به حساب می‌آورد و بازنمایی معنایی را تقویت می‌کند. مدل CPA-SA در برابر چندین مدل پیشرفته بر روی چهار مجموعه داده عمومی مانند توئیت‌های جریانات عمومی SemEval آزمایش شده و پیشرفت‌های قابل توجهی گزارش شده است. این مطالعه همچنین یک تابع زیان اصلاح‌شده برای حل مشکل عدم تعادل تعداد قطبیت‌ها در تحلیل احساسات پیشنهاد می‌کند. با این حال، در حالیکه به مشکل عدم تقارن تا حدی پرداخته شده، تابع وزنی تعیین موقعیت خاص جنبه به اندازه کافی برای زمینه‌های متنی پیچیده‌تر و تخصصی انعطاف‌پذیر نیست. همچنین، اتکای این مطالعه به ابرپارامترهای ثابت و مرزهای آستانه ممکن است سازگاری آن را در سناریوهای مختلف دنیای واقعی محدود کند.

یونژیا و همکاران[27] (۲۰۲۱)[۳۷]، روش جدیدی را برای تحلیل احساسات جنبه‌محور به نام شبکه حافظه زمینه و جنبه

[26] Embedding Vectors

[27] Yanxia et al.

(CAMN) پیشنهاد کرده‌اند. این روش از شبکه‌های حافظه عمیق، BiLSTM و مکانیسم‌های توجه چندسر برای گرفتن بهتر ویژگی‌های احساسی در متون، به ویژه متون کوتاه مانند توییت‌ها یا بررسی‌های محصول، استفاده کرده است.

شیائوفی و همکاران[28] (۲۰۲۱)[۴۱]، مدلی را پیشنهاد کردند که از وابستگی‌های متنی گلوبال و محلی برای کاوش در ساختار متن با استفاده از شبکه‌های کانولوشن گراف (GCNs) استفاده می‌کند. مدل GL-GCN از وابستگی نحوی و اطلاعات ترتیبی برای کاوش ساختارهای محلی استفاده کرده، در حالیکه یک نمودار سند-کلمه برای به تصویر کشیدن وابستگی‌های گلوبال کلمات در متن می‌سازد.

چوماکوف و همکاران (۲۰۲۳)[۴۲]، روشی برای استخراج سه‌گانه احساسات جنبه[29] (ASTE) با استفاده از GPT معرفی کردند. این رویکرد از تکنیک‌های یادگیری و تنظیم دقیق با مدل‌های GPT برای پردازش و تجزیه و تحلیل مؤثر داده‌های متنی در حوزه‌های مختلف، از جمله موارد ناشناخته، استفاده می‌کند. این روش دارای توانایی کاهش وابستگی به داده‌های گسترده دامنه خاص و حفظ عملکرد بالا در برنامه‌های کاربردی دنیای واقعی را دارا می‌باشد. در جدول (۱) خلاصه مقالات فوق به همراه نقاط ضعف آنها تشریح شده است.

جدول (۱): تحقیقات پیشین در تحلیل احساسات حنبه‌محور رمزارزها با مدل‌های پایه یادگیری عمیق و پیش‌آموزش‌دیده

مقاله	نوآوری	چالش تحقیقاتی
سیهان و همکاران (۲۰۲۳)	استفاده از معماری CNN-RNN با مکانیسم‌های توجه و بهبود گروهی برای تحلیل احساسات	پیچیدگی محاسباتی بالا و نیاز به داده‌های برچسب‌گذاری‌شده بیشتر برای آموزش بهتر
مودیج و	استفاده از تکنیک‌های تعبیه	نیاز به بررسی تعمیم‌پذیری

[28] Xiaofei et al.
[29] Aspect Sentiment Triplet Extraction (ASTE)

جدول (۱): تحقیقات پیشین در تحلیل احساسات حنبه‌محور رمزارزها با مدل‌های پایه یادگیری عمیق و پیش‌آموزش‌دیده

مقاله	نوآوری	چالش تحقیقاتی
همکاران (۲۰۲۳)	گراف برای استخراج روابط معنایی و نحوی در تحلیل احساسات	تکنیک تعبیه گراف به سایر زبان‌ها و دامنه‌ها
اسالم و همکاران (۲۰۲۳)	ارائه مدل ترکیبی LSTM-GRU برای تحلیل احساسات و تشخیص عواطف در توییت‌های ارزهای دیجیتال	تعادل‌بخشی به داده‌ها و نیاز به بهینه‌سازی بیشتر برای کاهش زمان محاسباتی
داوچف و همکاران (۲۰۲۰)	استفاده از مدل‌های انتقال یادگیری مانند روبرتا برای تحلیل احساسات و پیش‌بینی قیمت بیت‌کوین	چالش‌های مرتبط با عدم تعادل داده‌ها و نیاز به بهبود دقت پیش‌بینی با استفاده از داده‌های ترکیبی بیشتر
ژائو و همکاران (۲۰۲۰)	استفاده از GCN برای دریافت وابستگی‌های احساسات بین جنبه‌ای در جملات مختلف	تعمیم‌پذیری در دامنه‌های متفاوت و عدم بررسی استحکام مدل
زونیک و همکاران (۲۰۲۱)	استفاده از GCN بر روی نمودارهای وابستگی نحوی برای تحلیل احساسات جنبه‌محور	عدم تعادل طبقاتی در مجموعه دادگان، فاقد بحث مفصل در مورد کارایی محاسباتی
لی و همکاران (۲۰۲۱)	مدل برچسب‌گذاری فاصله محور (QDSL) برای شناسایی جفت‌های جنبه-نظر	عدم تعادل کلاس در مجموعه داده، تنظیم دقیق مدل پیشنهادی و آزمایش بر روی دادگان با زمینه متفاوت
هبت و همکاران (۲۰۲۲)	ترکیب GRU و CNN با XLNet برای تحلیل احساسات دادگان فرانسوی	عدم تنظیم مدل‌های عصبی عمیق نسبت به وظیفه و نوع داده ورودی
کبیر و همکاران (۲۰۲۳)	چارچوب ترکیبی شناسایی وابستگی نحوی با لغت‌نامه و رگرسیون لجستیک	ابهام در اثربخشی مدل پیشنهادی بر روی دامنه‌های مختلف
بوهوانگ و همکاران (۲۰۲۲)	مدل CPA-SA با استفاده از اطلاعات موقعیت خاص جنبه برای تحلیل احساسات جنبه‌محور	انعطاف‌پذیری ناکافی برای زمینه‌های متنی پیچیده‌تر و اتکا به ابرپارامترهای ثابت و مرزهای آستانه

جدول (۱): تحقیقات پیشین در تحلیل احساسات حنبه‌محور رمزارزها با مدل‌های پایه یادگیری عمیق و پیش‌آموزش‌دیده

مقاله	نوآوری	چالش تحقیقاتی
یونژیا و همکاران (۲۰۲۱)	شبکه حافظه زمینه و جنبه (CAMN) برای تحلیل احساسات جنبه‌محور	عدم تعادل کلاس‌های قطبیت و تنظیم ابرپارامترها
شیائوفی و همکاران (۲۰۲۱)	مدل GL-GCN با استفاده از وابستگی‌های متنی گلوبال و محلی برای تحلیل احساسات جنبه‌محور	استحکام مدل در سراسر حوزه‌ها و پیچیدگی محاسباتی بررسی نشده
چوماکوف و همکاران (۲۰۲۳)	استفاده از GPT برای استخراج سه‌گانه احساسات جنبه (ASTE)	احتمال بالقوه بیش‌برازش در سناریوهای یادگیری با شات محدود

در جدول (۲)، مدل پیشنهادی به صورت خلاصه و از نظر ساختار معماری و نوآوری با مقالات جدول (۱) مقایسه شده است.

جدول (۲): مقایسه تحقیقات پیشین در تحلیل احساسات رمزارزها با مدل پیشنهادی

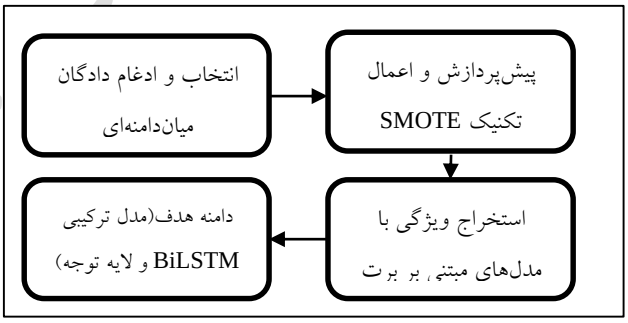
مقاله	مقایسه
سیهان و همکاران (۲۰۲۳)	مدل پیشنهادی با بهره‌گیری از یادگیری انتقالی شات صفر، قابلیت تعمیم‌پذیری بیشتری نسبت به LSTM دارد و می‌تواند با استفاده از تکنیک شات صفر کارایی بیشتری در دامنه‌های مختلف به داشته باشد.
مودیچ و همکاران (۲۰۲۳)	روش پیشنهادی با استفاده از مدل‌های عمیق و یادگیری انتقالی، در عین کاهش پیچیدگی محاسباتی، دقت بالاتری نسبت به تکنیک‌های تعبیه گراف ارائه می‌دهد و نیازی به برچسب‌گذاری دستی ندارد.
اسالم و همکاران (۲۰۲۳)	هر دو مقاله از یادگیری انتقالی شات صفر استفاده می‌کنند، اما مطالعه حاضر با بهره‌گیری از بهینه‌سازی داده‌ها، نرخ داده وفق‌پذیر و تنظیمات بهتر ابرپارامترها، می‌تواند دقت و کارایی بیشتری در تحلیل احساسات جنبه‌محور در حوزه رمزارزها ارائه می‌دهد.
داوچف و همکاران (۲۰۲۰)	روش پیشنهادی با استفاده از شبکه‌های عمیق و یادگیری شات صفر، تعمیم‌پذیری و دقت بالاتری نسبت به روش‌های مبتنی بر مدل‌های از پیش آموزش‌دیده بدون تکنیک شات صفر دارد. این رویکردها برای کارایی بهتر نیاز به داده‌های ترکیبی بیشتری در دامنه مبدا دارند.
ژائو و همکاران (۲۰۲۰)	مقاله حاضر با استفاده از یادگیری انتقالی شات صفر، قابلیت تعمیم‌پذیری بهتری نسبت به مدل GCN دارد و از آنجا که نیازی به داده‌های برچسب‌گذاری‌شده ندارد، در دامنه‌های متفاوت بهتر عمل می‌کند. همچنین در این پژوهش، بر روی استحکام مدل و تنظیم دقیق آن تمرکز بیشتری دارد.
زونیک و همکاران (۲۰۲۱)	روش پیشنهادی از روش‌های بهینه‌سازی و تعادل‌دهی داده‌ها استفاده می‌کند که مشکلات عدم تعادل طبقاتی را که در روش زونیک وجود دارد را مرتفع می‌کند. همچنین، کارایی محاسباتی مدل پیشنهادی به دلیل استفاده از مدل‌های سبک‌تر مانند دستیل‌برت بهبود یافته است.
لی و همکاران (۲۰۲۱)	در این پژوهش، از یادگیری انتقالی شات صفر استفاده شده که نیاز به تنظیم دقیق مدل و تعادل کلاس‌ها را کاهش می‌دهد. این روش همچنین در دامنه‌های مختلف بدون نیاز به داده‌های برچسب‌گذاری‌شده عملکرد بهتری ارائه می‌دهد.
هبت و همکاران (۲۰۲۲)	مقاله حاضر با استفاده از شبکه دیستیل‌برت و یادگیری انتقالی شات صفر، تنظیمات مدل را نسبت به وظیفه و نوع داده بهینه‌تر می‌کند.
کبیر و همکاران (۲۰۲۳)	روش پیشنهادی با استفاده از مدل‌های عمیق و یادگیری انتقالی، کارایی بیشتری در دامنه‌های مختلف دارد و وابستگی کمتری به تنظیمات دستی و لغت‌نامه‌های از پیش تعیین‌شده دارد.
بوهوانگ و همکاران (۲۰۲۲)	مقاله حاضر با انعطاف‌پذیری بیشتری در زمینه‌های متنی پیچیده عمل می‌کند و از ابرپارامترهای ثابت و مرزهای آستانه استفاده نمی‌کند، که باعث بهبود عملکرد در دامنه‌های مختلف می‌شود.
یونژیا و همکاران (۲۰۲۱)	مقاله حاضر از یادگیری انتقالی شات صفر استفاده می‌کند که مشکلات عدم تعادل کلاس‌های قطبیت و تنظیم ابرپارامترها را بهتر حل می‌کند و نیاز به داده‌های گسترده برچسب‌گذاری‌شده را کاهش می‌دهد.
شیائوفی و همکاران	این مقاله با تمرکز بر تعمیم‌پذیری و بهینه‌سازی مدل، استحکام بیشتری در دامنه‌های مختلف دارد و پیچیدگی محاسباتی کمتری

جدول (۲): مقایسه تحقیقات پیشین در تحلیل احساسات رمزارزها با مدل پیشنهادی

مقاله	مقایسه
چوماکوف و همکاران (۲۰۲۳)	نسبت به مدل GL-GCN دارد. مقاله حاضر از یادگیری انتقالی شات صفر استفاده می‌کند که احتمال پیش‌برازش را در سناریوهای یادگیری با شات محدود کاهش می‌دهد و نیازی به تنظیمات پیچیده مدل GPT ندارد.

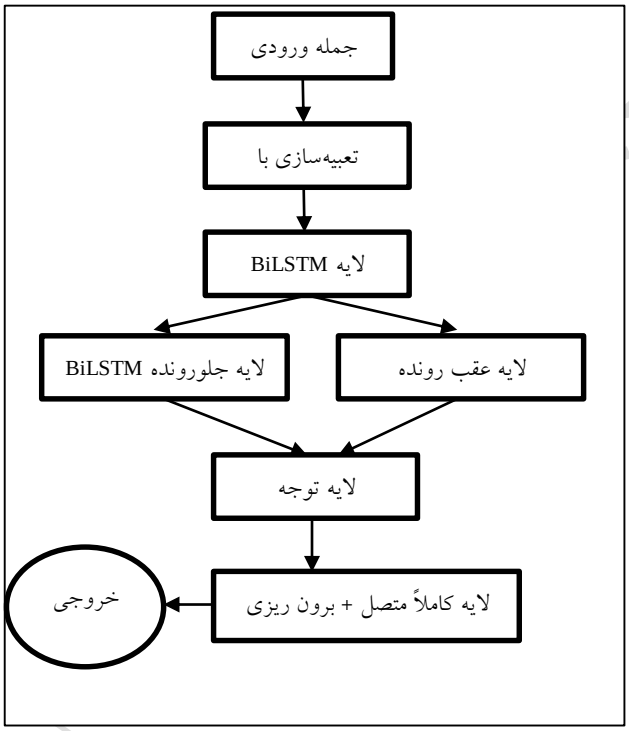
۳. روش پیشنهادی

مدل‌های پیشنهادی در این پژوهش، ترکیبی از شبکه‌های پیش‌آموزش مبتنی بر برت و شبکه عصبی عمیق BiLSTM می‌باشد. چارچوب کلی این‌مدل‌ها در شکل (۲) نشان داده شده است.



شکل ۲- مدل عمومی شبکه‌های عصبی عمیق پیش‌آموزش دیده ترکیبی برای تحلیل احساسات رمزارزها

همانگونه که در شکل (۲) مشاهده می‌شود در اینجا برای بهبود دقت و کارایی مدل تحلیل احساسات جنبه‌محور، از ترکیب شبکه عصبی BiLSTM با یک لایه توجه استفاده شده است. ساختار BiLSTM به‌طور همزمان اطلاعات متنی را از دو جهت (جلو به عقب و عقب به جلو) پردازش می‌کند، که این ویژگی به مدل اجازه می‌دهد تا وابستگی‌های طولانی‌مدت را در هر دو جهت شناسایی کند. این قابلیت باعث می‌شود که مدل بتواند اطلاعات بیشتری از توالی جملات استخراج کرده و درک بهتری از ساختار معنایی متن‌ها داشته باشد. پس از عبور داده‌های متنی از لایه BiLSTM، از لایه توجه برای تقویت بخش‌های مهم‌تر جملات استفاده می‌شود. به بیان دیگر، خروجی نهایی هر طرف BiLSTM به لایه توجه داده شده سپس، این بردارهای خروجی در کنار هم قرار می‌گیرد و مکانیزم توجه روی هر توکن خروجی BiLSTM اعمال شده و سبب غنی‌تر شدن توکن‌ها می‌شود. بنابراین، لایه توجه به مدل امکان می‌دهد که روی کلمات یا عبارات کلیدی که تأثیر بیشتری بر روی تحلیل احساسات دارند، تمرکز بیشتری داشته باشد. این امر به بهبود دقت در شناسایی احساسات مرتبط با جنبه‌های خاص جملات کمک می‌کند. خروجی لایه توجه سپس به لایه‌ای کاملاً متصل تزریق می‌شود تا فرایند دسته‌بندی نهایی انجام گیرد و احساسات مرتبط با جنبه به یکی از قطبیت‌های مثبت، منفی یا خنثی تخصیص داده شود. فرایند عملکرد BiLSTM و لایه توجه در شکل (۳) نشان داده شده است.



شکل ۳- فرایند پردازش جملات در BiLSTM و لایه توجه

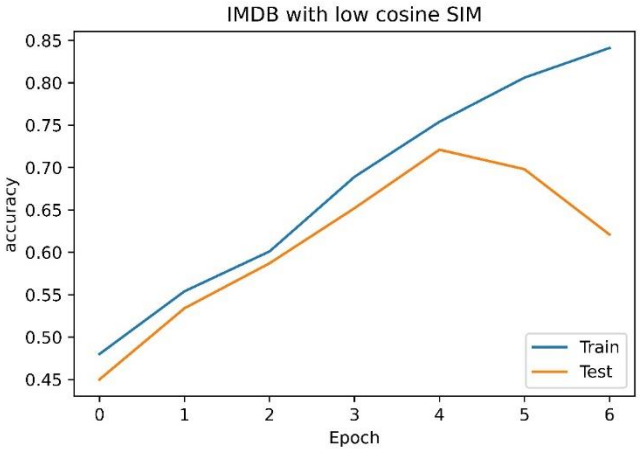
در شکل (۳) فرایند پردازش جملات بعد از ورود به لایه‌های BiLSTM و توجه نشان داده شده است. همچنین در این پژوهش، در ابتدا شباهت دادگان انتخاب شده جهت آموزش

شبکه عصبی پیش‌آموزش‌دیده با دادگان رمزارزها مقایسه می‌شود. برای شباهت سنجی از تابع شباهت پیرسون بر اساس شبه کد (۱) استفاده شده است. سپس دادگانی انتخاب می‌گردند که دارای شباهت کسینوسی بیشتر از ۴۰ درصد با دادگان دامنه هدف باشند.

شبه کد (۱): انتخاب دادگان دامنه هدف با معیار شباهت پیرسون

```
Import Pearson similarity
Res= [], Datasets= {IMDB, Crypto, Sentiment}, DatasetT= Tweet_Crypto
for (i=1 to #Datasets): {
    Res[i] = Pea_Sim (Datasetsi[tweet], Tweet_Crypto]
    If Res[i]>0.4 then:
      Datasetsi[flag]= True, Else: Datasetsi[flag]= False
End For
Datasets-Final= [x in Datasets[tweets] if Datasets[flag]==True]
```

نتایج شکل نمودار (۴) نشان می‌دهد، در صورتیکه مدل با دیتاست‌های که دارای شباهت پیرسون کمتر ۴۰ درصد باشند آموزش داده شود، انتقال دانش می‌تواند تاثیر منفی بر کارایی دامنه هدف داشته باشد.



شکل نمودار ۴- بروز انتقال منفی با استفاده از دیتاست IMDB در دامنه مبدا

نمودار شکل (۴)، نشان دهنده بروز حالت انتقال منفی و بیش‌برازش شبکه عصبی عمیق می‌باشد. در اینجا افت کارایی شبکه به دلیل انتقال منفی دانش می‌باشد زیرا، مدل برت پایه که از دادگان احساسات توئیت‌های بیت‌کوین به عنوان دادگان مبدا استفاده کرده، دقت ۸۵٪ بعد از ۶ دور دست یافته است. پس می‌توان نتیجه گرفت، کاهش کارایی مدل به دلیل پارامترهای شبکه نبوده است. بنابراین انتخاب مجموعه دادگان غنی در دامنه مبدا اگرچه می‌تواند سبب افزایش کارایی مدل در دامنه هدف گردد، با این حال، انتخاب مجموعه دادگان نامشابه یا با شباهت کمتر می‌تواند سبب انتقال منفی دانش شود [۱۵].

در ادامه مجموعه دادگان منتخب جهت استفاده در شبکه‌های پیش‌آموزش‌دیده دامنه مبدا، به یکدیگر الحاق می‌گردد. سپس عملیات پیش پردازش متن روی آنها انجام می‌شود. این عملیات شامل نشانه‌گذاری، حذف کلمات توقف، بن‌سازی، یکسازی‌سازی متن و حذف کارکترهای زائد مانند لینک‌ها می‌باشد. سپس داده‌های پاکسازی شده با تکنیک SMOTE متوازن شده و در نهایت به شبکه عصبی عمیق پیش‌آموزش‌دیده دیستیل‌برت در دامنه مبدا تزریق می‌شود.

SMOTE تکنیکی آماری برای متعادل‌سازی دادگان است که سبب می‌شود حساسیت مدل به کلاسی خاص در دادگان کم شود. به بیان رسمی، برای یک نمونه کلاس اقلیت x، SMOTE یک نمونه تصادفی $\bar{x}$ را از k نزدیکترین همسایه x انتخاب می کند. سپس نمونه‌ای جدید به اسم x_{new}با درون‌یابی بین x و $\bar{x}$ توسط رابطه (۳) ایجاد می‌شود.

$$x_{new} = x + \lambda * (\acute{x} - x) \quad (۳)$$

در رابطه (۳)، λ عددی تصادفی بین ۰ و ۱ است. این فرآیند تا زمانی که کلاس‌ها متعادل شوند تکرار می‌شود. SMOTE منجر به ارائه مجموعه دادگانی متعادل‌تر می‌شود که برای آموزش مدل‌های یادگیری ماشین بسیار مهم است. زیرا، داده‌های نامتعادل می‌تواند منجر به مدل‌های مغرضانه شود که عملکرد ضعیفی در کلاس اقلیت دارند. این تکنیک توانایی مدل را برای تعمیم و عمل موثر بر روی داده های دیده نشده افزایش می‌دهد. همچنین در این پژوهش برای انتخاب ابر پارامترهای مدل عصبی عمیق پایه BiLSTM در دامنه هدف، از روش گریدی استفاده

شده است. روش گریدی این امکان را می‌دهد که مجموعه‌ای از بهترین ابرپارامترها برای شبکه عصبی عمیق با مانتیور کردن معیارهای کارایی شبکه انتخاب شود. معیاری که معمولاً در مقالات برای مانتیور کردن انتخاب می‌شود، تابع نرخ خطا می‌باشد. در این مقاله، نرخ خطای داده اعتبار سنجی برای مانتیور کردن انتخاب شده است. در شبه کد (۲) نحوه انتخاب ابرپارامترها نشان داده شده است.

شبه کد (۲): استفاده از رویکرد گریدی برای انتخاب ابرپارامترهای شبکه عصبی عمیق

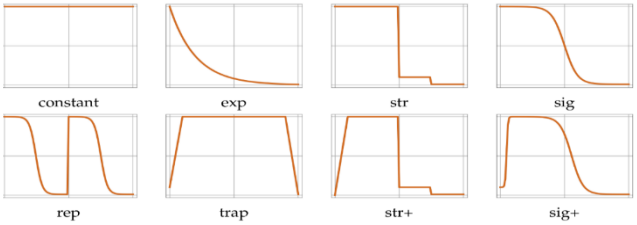
```
Optimizer ={'SGD','Adam','AdamW'}
Dropout = {0.1,0.2,0.3,0.4}
Batch Size = {16,32,64,128}
Def grid_search (Optimizer, Dropout, Batch Size) {
    For (i=0 to Number_of_Epochs) {
      Calc_Lossi = (Optimizeri, Dropouti, Batch Sizei)}
Return (Min (Calc_Loss))
End Def
```

در این پژوهش سه ابرپارامتر تابع بهینه‌سازی، نرخ برون‌ریزی شبکه و طول بچ با استفاده از روش گریدی تعیین شده است. این سه ابر پارامتر تاثیر بسیار زیادی بر کارایی شبکه‌های عصبی دارند [۴۳]. نتایج آزمایشات نشان داده است انتخاب سه ابرپارامتر به شرح جدول (۳) بهترین کارایی را داشته است. از این رو برای پیاده‌سازی کلیه مدل‌ها در این پژوهش از مقادیر ابرپارامترهای جدول (۳) استفاده می‌شود.

جدول (۳): ابر پارامترهای منتخب

طول بچ	۶۴
تابع بهینه‌ساز	*AdamW*
نرخ برون ریز	۰,۲

همچنین، نرخ یادگیری شبکه عصبی در این پژوهش برخلاف اغلب مقالات [۴] به صورت پویا و با استفاده تکنیک وارم‌آپ[30] تعیین می‌گردد. تکنیک وارم‌آپ [۴۴]، رویکردی نوین است که از اندازه گام در شیب نزولی مشتق‌های گرادیان را در ابتدای آموزش شبکه کوچک انتخاب می‌کند. در ادامه، نرخ یادگیری به صورت خطی یا غیرخطی در چند دوره افزایش یافته و سپس به سمت صفر میل می‌کند. مدل‌های یادگیری ماشین در ابتدای آموزش از مدل ایده‌آل فاصله زیادی دارد. بنابراین، نرخ یادگیری بیش از حد بزرگ باعث بی‌ثباتی عددی می‌شود. در نتیجه، آموزش شبکه عصبی با دقت بالا در دورهای اول این امکان را می‌دهد در مراحل میانی نرخ یادگیری افزایش یافته در نتیجه همگرایی شبکه تسریع گردد [۴۴]. در شکل (۵) متدهای مختلف وارم‌آپ را نشان داده شده است.



شکل ۵- متدهای مختلف وارم‌آپ [۴۴]

در این پژوهش از روش +sig استفاده شده است. تکنیک وارم‌آپ در واقع نوعی روش تبرید فلزات[31] است که در ابتدا نرخ یادگیری به سمت صعودی میل کرده (افزایش دما)، در مراحل میانی این نرخ ثابت نگه داشته شده و سپس به سمت صفر (انجماد) میل می‌کند. در این پژوهش همچنین دو تکنیک تنظیم دقیق برای شبکه‌های پیش‌آموزش دیده مبتنی بر برت یعنی دستیل‌برت و XLNet با یکدیگر مقایسه شده است. بطور کلی سه روش برای تنظیم شبکه‌های عصبی پیش‌آموزش‌دیده به شرح زیر وجود دارد [۵]:

۱- آموزش کل شبکه: می‌توان کل شبکه پیش‌آموزش‌دیده را با دادگان دامنه هدف آموزش داد سپس خروجی را به لایه‌ای متراکم با فعالسازهای سیگموید یا سافت‌مکس متصل کرد. در این حالت، خطا در شبکه

[30] Warmup

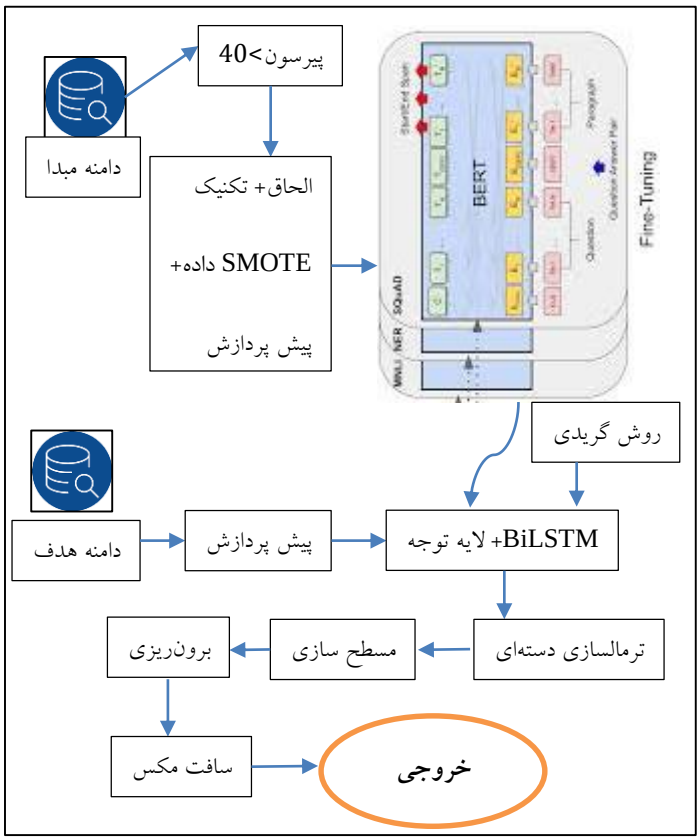
[31] Annealing

منتشر شده و وزن‌های شبکه بر اساس مجموعه داده دامنه هدف بروز می‌شود.

۲- انجماد برخی از لایه‌ها : روش دیگر تنظیم مدل‌های پیش‌آموزش‌دیده، آموزش جزئی آنها است. می‌توان وزن لایه‌های اولیه را ثابت نگه داریم و فقط لایه‌های بالاتر را دوباره آموزش دهیم. این روش پیشرفته‌تر است و می‌توان آزمایش کرد بسته به غنای دادگان دامنه هدف، چند لایه باید منجمد شود و چند لایه باید آموزش داده شود.

۳- کل شبکه را منجمد کنیم: می‌توان تمام لایه‌های شبکه را منجمد کرد و چند لایه شبکه عصبی (پایه یا عمیق) به آن متصل کنیم و تنها شبکه پایانی را دوباره آموزش دهیم. در این روش تنها وزن لایه‌های نهایی پیوست شده در فرایند آموزش بروز رسانی می‌شود.

در این پژوهش دو روش تنظیم دقیق آموزش و انجماد کل شبکه با یکدیگر مقایسه شده است. چارچوب کلی مدل پیشنهادی در شکل (۶) نشان داده شده است.



شکل ۶- معماری پیشنهادی یادگیری انتقالی عمیق مبتنی بر شبکه‌های تنظیم دقیق پیش‌آموزش دیده برت

همانگونه که در شکل (۶) نشان داده شده است، در ابتدا دادگان دامنه هدف به وسیله معیار شباهت پیرسون انتخاب می‌شوند. سپس، توئیت و برچسب‌های دادگان منتخب که دارای شباهت بیشتر از ۴۰ درصد با دادگان رمزرازها هستند، با یکدیگر الحاق شده و دادگان دامنه مبدا را تشکیل می‌دهند. در ادامه برخی از عملیات پیش‌پردازش معمول در پردازش زبان طبیعی مانند ریشه‌یابی، حذف کلمات توقف، یکسان‌سازی حروف، حذف عبارات زائد مانند لینک‌ها و نشانه‌گذاری بر روی آنها انجام شده و بعد از اعمال تکنیک متعادل‌سازی داده با SMOTE به شبکه پیش‌آموزش‌دیده از خانواده برت تزریق می‌شود. ذکر این نکته ضروری است که برای نشانه‌گذاری هم در دامنه مبدا و هم در دامنه هدف از نشانه‌گذار مبتنی بر برت استفاده می‌شود. برای مثال اگر شبکه پیش‌آموزش‌دیده، XLNet باشد از نشانه‌گذار XLNet استفاده می‌شود. در ادامه دادگان دامنه مبدا

برای استخراج دانش به یکی از انواع برت تزریق می‌گردد. در این پژوهش از سه نسخه برت پایه، دستیل‌برت و XLNet استفاده شده است. سپس برای انتخاب ابرپارامترهای شبکه‌پیش‌آموزش‌دیده از تکنیک گریدی (شبه کد ۲) استفاده شده است. ویژگی‌های استخراج شده از دامنه مبدا با دو روش تنظیم به شبکه BiLSTM با ترکیب لایه توجه در دامنه هدف تزریق می‌شود. در این پژوهش برای مقایسه و پاسخ به سوالات پژوهشی از دو روش تنظیم آموزش کل شبکه و انجمادکل وزن‌های شبکه دامنه مبدا، استفاده شده است.

در دامنه هدف نیز بعد از پیش‌پردازش دادگان رمزارز، داده‌ها نشانه‌گذاری‌شده و به شبکه عصبی عمیق BiLSTM تزریق شده است. سپس ویژگی‌های استخراج شده از شبکه‌پیش‌آموزش‌دیده دامنه هدف و دامنه مبدا با یکدیگر الحاق می‌شود. وجود لایه الحاق سبب می‌شود شبکه‌های که از نظر ساختار (برای مثال تعداد لایه‌های پنهان) با یکدگیر یکسان نیستند بتوانند به معماری یکسانی دست یابند. همچنین لایه الحاق در این پژوهش می‌تواند مانند یک سوئیچ عمل کند، یعنی هرگاه نیاز به دانش دامنه هدف نبود و یا شبکه پیش‌آموزش‌دیده تغییر کرد می‌توان این تغییرات را بدون دستکاری دامنه هدف اعمال کرد. این امر بیشتر در موارد مقایسه شبکه‌های عصبی پایه و پیش‌آموزش دیده کاربرد دارد. برای مثال در مواردی که می‌خواهیم کارایی شبکه برت+BiLSTM را با BiLSTM مقایسه کنیم به راحتی می‌توان بدون کدنویسی مجدد داده‌های منتقل شده از برت را نادیده گرفت. در این پژوهش کارایی شبکه‌های عصبی عمیق پایه، پیش‌آموزش‌دیده و ترکیبی براساس معماری ارائه شده در شکل (۴) با یکدیگر مقایسه شده است. شبکه‌های مورد مطالعه این پژوهش در جدول (۴) ذکر شده است.

جدول (۴): معماری‌های عصبی عمیق پیاده‌سازی شده

نام مدل	نوع مدل
LSTM BiLSTM	شبکه عصبی عمیق پایه
CNN برت پایه	شبکه عمیق پیش‌آموزش‌دیده
CNN+BiLSTM	شبکه عصبی عمیق پایه ترکیبی
BiLSTM+ برت(پایه، دستیل)	شبکه عمیق پیش‌آموزش‌دیده ترکیبی

۴. نتایج آزمایشات تجربی

در این بخش در ابتدا معیارهای ارزیابی شبکه‌های عصبی عمیق برای تحلیل احساسات جنبه‌محور رمزارزها را توضیح داده شده؛ سپس، دادگان مورد استفاده در دامنه هدف و مبدا مورد بررسی قرار گرفته و در نهایت نتایج آزمایشات تجربی تشریح می‌شود.

۱،۴. معیارهای ارزیابی

اولین معیار مرسوم در ارزیابی روش‌های طبقه‌بندی متن مانند تحلیل احساسات معیار صحت می‌باشد [۴۵]. این معیار با تقسیم تعداد پیش‌بینی‌های صحیح بر تعداد کل پیش‌بینی‌ها محاسبه می‌شود. معیار صحت از طریق رابطه (۴) بدست می‌آید.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۴)$$

در رابطه (۴)، TP تعداد مثبت واقعی، TN منفی واقعی، FP مثبت کاذب و FN تعداد منفی کاذب پیش‌بینی شده می‌باشد. یکی دیگر از معیارهای مرسوم، معیار دقت[32] است. این معیار برای گزارش تعداد کل پیش‌بینی‌های صحیح شبکه می‌باشد. این معیار با تقسیم تعداد پیش‌بینی‌های مثبت توسط شبکه بر تعداد کل برچسب‌های مثبت در دادگان برای هر قطبیت مثبت، منفی و خنثی بدست می‌آید. معیار دقت به صورت رابطه (۵) تعریف می‌شود.

$$Precision = \frac{TP}{TP + FP} \quad (۵)$$

[32] Precision

معیار بعدی فراخوان[33] است. فراخوان نشان‌دهنده تمام برچسب‌های مثبتی است که توسط شبکه به درستی پیش‌بینی شده‌اند. این معیار با تقسیم برچسب‌های مثبت پیش‌بینی‌شده توسط شبکه با مجموع تعداد کل برچسب‌های مثبت و تعداد منفی‌های کاذب پیش‌بینی شده بدست می‌آید. معیار فراخوان به صورت رابطه (۶) تعریف می‌شود.

$$Recall = \frac{TP}{TP + FN} \quad (۶)$$

معیار دیگر F1-Score است. این معیار ترکیبی از دقت و فراخوان می‌باشد و به صورت رابطه (۷) تعریف می‌گردد.

$$F1 - SCORE = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \quad (۷)$$

معیار بعدی، تابع ضرر شبکه می‌باشد که آنتروپی متقابل طبقه‌بندی[34] نام دارد. هر میزان این معیار به سمت صفر میل‌کند، نشان می‌دهد شبکه از کارایی بالاتری برخوردار است. به این معیار خطای سافت‌مکس نیز می‌گویند و از طریق رابطه (۸) بدست می‌آید.

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}, CE = -\sum_i^C t_i log(f(s)_i) \quad (۸)$$

در رابطه (۸)، S امتیاز اکتسابی توسط تابع سافت‌مکس (f) برای کلاس i‌ام است. در این رابطه، تابع ضرر آنتروپی متقابل طبقه‌بندی را با CE نشان داده که در آن C نشان دهنده تعداد کلاس‌ها و t نشان دهنده برچسب واقعی کلاس i‌ام می‌باشد. آخرین معیار، نمودار منحنی AUC-ROC است که نشان دهنده دقت مدل در تشخیص هر کلاس قطبیت می‌باشد. ای معیار در صورت رابطه (۹) تعریف شده است.

$$ROC - AUC = \int TP(FP)\, dFP \quad (۹)$$

۲،۴. معرفی دادگان مورد ارزیابی

در این پژوهش از سه مجموعه دادگان مورد بررسی جهت استفاده در دامنه مبدا، در جدول (۵) تشریح شده‌اند.

جدول (۵): دادگان دامنه مبدا

عنوان	مثبت	منفی	خنثی	مجموع
IMDB	۱۲،۵۰۰	۱۲،۵۰۰	۰	۵۰،۰۰۰
تحلیل احساسات توییتر	۲۲،۸۰۷	۲۱،۱۰۹	۱۸،۳۰۶	۶۲،۲۲۲
احساسات توئیت‌های بیت‌کوین	۲۲،۹۳۷	۵،۹۸۳	۲۱،۹۳۲	۵۰،۸۵۲

در زیر این دادگان معرفی شده است.

۱- دادگان IMDB: مجموعه داده بزرگ نقد فیلم‌های سایت IMDB، برای طبقه‌بندی احساسات دو یا چند کلاسه پیشنهاد شده است. این مجموعه دادگان شامل مجموعه‌های آموزشی و ارزیابی برچسب‌دار می‌باشد. یکی از ویرایش‌های دو کلاسه این دادگان، شامل ۲۵،۰۰۰ هزار نمونه آموزشی و ارزیابی است. این ویرایش، دارای ۱۲،۵۰۰ برچسب مثبت و منفی برای داده‌های ارزیابی و آموزشی می‌باشد. این دادگان به دلیل شباهت پائین‌تر از ۴۰ درصد به عنوان داده میان‌دامنه‌ای برای آموزش مدل استفاده نشده است.

۲- دادگان تحلیل احساسات توییتر: این مجموعه داده شامل بیش از ۶۲ هزار توییت است که با استفاده از Api توییتر در سال ۲۰۲۲ استخراج شده است. توییت‌ها دارای برچسب مثبت، منفی و خنثی می‌باشد و از آن می‌توان در وظیفه تحلیل احساسات استفاده کرد. این مجموعه دادگان به‌طور مستقیم با موضوع

[33] Recall
[34] Categorical cross-entropy

تحقیق در ارتباط است زیرا همانند داده‌های اصلی مقاله شامل توییت‌های کاربران است. بنابراین، به بهبود دقت مدل در تحلیل احساسات توییت‌های مربوط به رمزارزها کمک می‌کند. در این دادگان متوسط طول جملات ۱۸,۹۹ کلمه و حداکثر طول بزرگترین جمله ۱۹۸ کلمه است. همچنین انحراف معیار برابر ۱۴,۳ و تعداد کلمات یکتا برابر ۳۴ هزار کلمه می‌باشد. این مجموع دادگان به مدل کمک می‌کند با توئیت‌ها تقریباً بزرگ نیز آشنا شود.

۳- دادگان توئیت‌های احساسات بیت‌کوین: این دادگان شامل توئیت با هشتگ بیت‌کوین بوده که از آپریل تا سپتامبر سال ۲۰۱۸ جمع‌آوری شده است. این مجموع دادگان شامل بیش از ۵۰ هزار توئیت دارای برچسب مثبت، منفی و خنثی می‌باشد. در این دادگان متوسط طول جملات ۱۷,۱۶ و حداکثر طول بزرگترین جمله ۳۹ کلمه است. همچنین انحراف معیار برابر ۱۳,۲ و تعداد کلمات یکتا تقریباً برابر ۵۵ هزار کلمه می‌باشد. این دادگان به دلیل ارتباط مستقیم با موضوع اصلی تحقیق، یعنی تحلیل احساسات جنبه‌محور در حوزه رمزارزها، انتخاب شده و شامل توییت‌هایی است که نظرات کاربران درباره بیت‌کوین را جمع‌آوری و برچسب‌گذاری کرده است. این مجموعه دادگان تاکنون در مقالات متعددی استفاده شده است [۴۶-۴۸].

در دامنه هدف نیز از مجموعه دادگان رمزارزها جمع‌آوری شده از توئیتر استفاده شده است [۶]. مجموعه دادگان دامنه هدف مربوط به نظرات بیش از ۵۰ تاثیرگذار (شخص یا شرکت) در مورد بیش از ۴۰ رمزارز است. این مجموعه داده از طریق API توییتر به مدت هشت ماه از سال ۲۰۲۱ تا ۲۰۲۲ جمع‌آوری شده است. هدف از این مجموعه داده جمع‌آوری توییت‌های تخصصی در مورد رمزارزهای مختلف است. در مجموعه داده‌های موجود در زمینه رمزارزها، تخصص کاربران توییت‌زننده نادیده گرفته شده و توییت‌ها تنها با استفاده از هشتگ استخراج شده‌اند. لذا، مجموعه داده استفاده شده در این مقاله کاربردهای زیادی در یادگیری ماشین و متن کاوی دارد. محققان می‌توانند از این مجموعه داده در تجزیه و تحلیل احساسات، یادگیری انتقالی در تحلیل احساسات و طبقه‌بندی متن استفاده کنند. همچنین، به دلیل ذخیره توییت‌های مربوط به هر رمزارز در جداولی جداگانه، محققان می‌توانند از این مجموعه داده در تحلیل احساسات مبتنی بر جنبه نیز استفاده کنند. یکی دیگر از کاربردهای این مجموعه داده، بررسی تاثیر توییت‌های تاثیرگذاران متخصص بر روند بازار رمزارزها است. زیرا بر اساس علم اقتصاد رفتاری [۳۳]، نظرات معامله‌گران بر روند بازار سهام تأثیرگذار است. این دیتاست شامل ۸۰ هزار سطر می‌باشد که به ترتیب دارای ۳۹،۵۲۶ هزار برچسب مثبت، ۲۸،۷۱۵ برچسب منفی و ۱۱،۷۵۹ بر چسب خنثی است. در هر دو مجموعه دادگان در دامنه مبدا و هدف، ۸۰ درصد برای آموزش شبکه و ۲۰ درصد برای ارزیابی استفاده شده است. همچنین، برای ارزیابی مدل مانند هوانگ و همکاران [۴۰] از مجموع دادگان SemEval که شامل توئیت‌های عمومی در حوزه‌های مختلف مانند رستوران و لپ‌تاپ است استفاده شده است. دیتاست‌های مورد ارزیابی در این پژوهش SemEval-2015 Task 12 [۴۹] و SemEval-2016Task 5 [۵۰] است که از این پس به صورت اختصاری Res2015 و Res2016 نامگذاری می‌شود. در جدول (۶) این مجموعه دادگان معرفی شده است.

جدول (۶): تنظیمات شبکه‌های عصبی عمیق

دادگان	آموزش	اعتبارسنجی	ارزیابی

	خنثی	منفی	مثبت	خنثی	منفی	مثبت	خنثی	منفی	مثبت
Res2015	۱۵۹	۲۵	۳۴۰	۴۴	۵	۱۴۷	۲۲۸	۲۹	۸۰۸
Res2016	۱۲۷	۲۹	۴۷۴	۶۰	۹	۱۹۱	۴۰۶	۴۵۴	۷۰۶

۳,۴. نتایج آزمایشات مدل‌های عصبی عمیق پایه، پیش‌آموزش‌دیده و ترکیبی

در ابتدا شباهت محتوایی داده‌های جمع‌آوری شده رمزارزها با سه مجموعه دادگان دامنه مبدا، ارزیابی شده که کمترین شباهت برابر با ۳۶,۵ درصد مربوط به مجموع دادگان IMDB می‌باشد. بنابراین این مجموعه داده به دلیل شباهت کمتر از ۴۰ درصد به عنوان داده میان‌دامنه‌ای استفاده نمی‌شود. در جدول (۷) نتایج این ارزیابی نشان داده شده است.

جدول (۷): محاسبه معیار شباهت کسینوسی

امتیاز	دادگان
۵۳,۲٪	Pear (sentiment, My_Crypto)
٪۶۴,۱	Pear (Crypto Tweets,My_Crypto)
٪۳۶,۵	Pear (IMDB,My_Crypto)

در این پژوهش برای مدل‌های ترکیبی پیش‌آموزش‌دیده مبتنی بر برت (XlNet,)، پایه و دستیل)، شبکه‌های عصبی عمیق پایه (CNN,BiLSTM) و شبکه‌های عصبی عمیق پایه ترکیبی (CNN+BiLSTM) از تنظیمات ذکر شده در جدول (۸) استفاده شده است.

جدول (۸): تنظیمات شبکه‌های عصبی عمیق

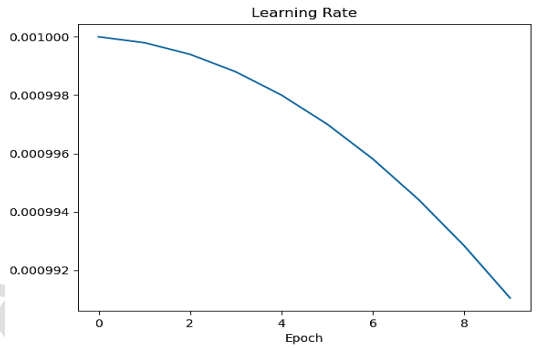
مقدار	نام پارامتر
۱۰	تعداد دور
AdamW	تابع بهینه‌ساز
وفقی (Warmup)	نرخ یادگیری
Model 1= True Model 2 = False	انجماد لایه‌ها
Softmax/ Sigmoid	توابع فعالساز
۳۲،۶۴،۱۲۸	طول بچ
۳۲،۶۴	تعداد لایه‌های مخفی BiLSTM
۰,۲	نرخ برون‌ریزی
۲×۲،۳×۳،۵×۵	طول کرنل CNN
آنتروپی متقابل طبقه‌بندی	تابع ضرر

تنها تفاوت در پیاده‌سازی شبکه‌ها این است که شبکه‌های عصبی عمیق پایه و پایه-ترکیبی تنظیمات مرتبط با انجماد لایه‌ها وجود ندارد. همچنین تنظیمات مرتبط با CNN مربوط به معماری ترکیبی CNN+BiLSTM است. روش پیشنهادی، بر روی سخت افزار با مشخصات جدول (۹)، پیاده‌سازی شده است.

جدول (۹): سخت افزار مورد استفاده

مشخصات	سخت افزار
Colab Google	محیط برنامه نویسی
16.0 GB	حافظه داخلی
پایتون ۳	زبان برنامه‌نویسی
Telsa T4	پردازنده گرافیکی
۱۱,۲	نسخه کودا

همانگونه که ذکر شد در این پژوهش از نرخ یادگیری وفق‌پذیر وارم‌آپ استفاده شده است. در نمودار شکل (۷)، میزان کاهش نرخ یادگیری در ده دور نشان داده شده است.



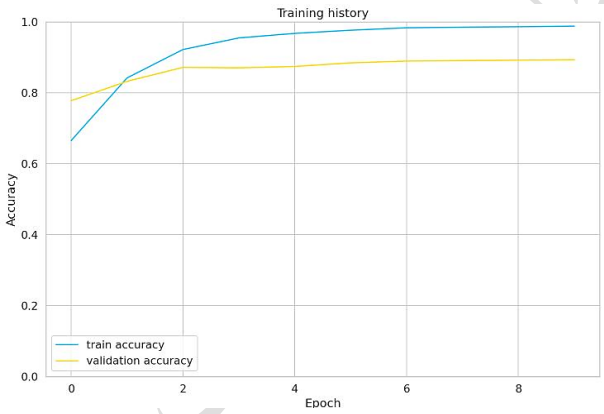
شکل ۷- تغییرات نرخ یادگیری

نتایج جدول (۱۰)، نشان دهنده کارایی مدل‌های XLNeT +RNN، BiLSTM+ دستیل برت (با انجماد و بدون انجماد لایه‌ها)، برت پایه و برت+ BiLSTM می‌باشد.

جدول (۱۰): دقت مدل‌های پیش‌آموزش‌دیده ساده و ترکیبی مبتنی بر برت بر روی دادگان حساسات توئیت‌های بیت‌کوین

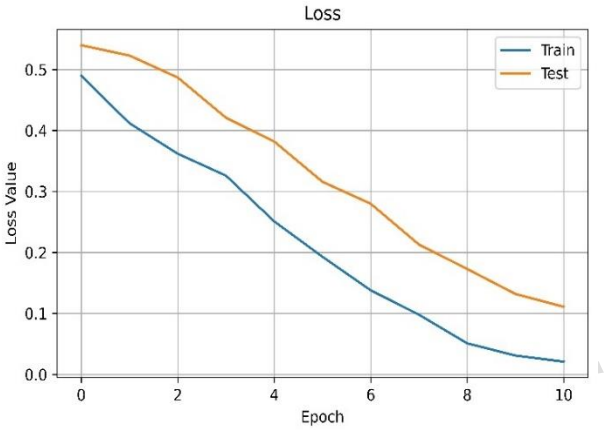
نام مدل	صحت	دقت	فراخوان	F1-Score
LSTM	۰,۶۵۱	۰,۶۲۳	۰,۶۴۳	۰,۶۴۶
BiLSTM	۰,۷۲۳	۰,۶۵۱	۰,۶۹۲	۰,۷۱۱
CNN+ BiLSTM	۰,۷۸۱	۰,۶۹۵	۰,۷	۰,۷۳۸
XLNeT +CNN	۰,۸۶۹	۰,۸۴۶	۰,۸۳۲	۰,۸۵
	۰,۷۵۱	انتقال منفی(با IMDB)		
BiLSTM+ distilbert (freeze)	۰,۸۷۶	۰,۸۴۶	۰,۸۲۵	۰,۸۴۶
BiLSTM +distilbert (unfreez)	۰,۸۸۲	۰,۸۵۷	۰,۸۴۱	۰,۸۶۱
برت پایه	۰,۸۰۶	۰,۷۸۶	۰,۷۷۴	۰,۷۸
برت پایه+ BiLSTM	۰,۸۱۷	۰,۷۹۲	۰,۷۸۵	۰,۸۰

همانگونه که در جدول (۱۰)، نشان داده شده، استفاده از دادگان IMDB به دلیل عدم شباهت دادگان مبدا به هدف، سبب بروز پدیده یادگیری انتقالی منفی شده است. همچنین، مدل‌های پیش‌آموزش‌دیده برت منجمد و بدون انجماد، تقریباً به نتایج مشابهی دست یافته‌اند. دلیل این امر، می‌تواند شباهت متن‌های مجموع دادگان دامنه مبدا و هدف باشد. دلیل دیگر، استفاده از مدل BiLSTM است. این مدل به دلیل دو جهته بودن و دسترسی به توالی‌های قبلی و بعدی [۵۱] برای بالابردن کارایی در وظیفه تحلیل احساسات استفاده شده است. با این حال، در کاربردهای دنیای واقعی آموزش کل لایه‌های شبکه‌های عصبی عمیق بسیار هزینه‌بر می‌باشد. بنابراین می‌توان با تحمل درصد ناچیزی از خطا از تکنیک‌های تنظیم دقیق مبنی بر انجماد لایه‌های استخراج‌گر ویژگی استفاده کرد. همچنین، می‌توان به لایه‌های که باید منجمد شود به صورت یک ابرپارامتر نگاه کرد. به عبارت دیگر، با در نظر گرفتن داده‌های دامنه هدف به صورت وفق‌پذیر برخی لایه‌ها را منجمد و برخی دیگر را آموزش داد. این امر سبب ایجاد موازنه بین سرعت آموزش و دقت شبکه‌های عصبی عمیق می‌گردد که در کارهای آینده مورد توجه قرار می‌گیرد. از سوی دیگر، می‌توان متصور بود در شبکه XLNeT نیز اگر شبکه عصبی بازگشتی با BiLSTM جایگزین شود به نتایج بهتری دست بیابیم. زیرا XLNeT مشکل فرض استقلال در برت را حل کرده و همچنین مدل BiLSTM دارای مشکل وابستگی طولانی مدت مانند RNN نیست. فرض استقلال، تصور می‌کند هر توکن پوشانده‌شده به تمام توکن‌های ورودی وابسته است اما از سایر توکن‌های پوشانده‌شده مستقل می‌باشد. همچنین در موارد ترکیب شبکه‌های عصبی عمیق پایه می‌بینیم، شبکه عصبی ترکیبی CNN+BiLSTM بهترین عملکرد را نسب به سایر شبکه‌ها بدست آورده است. این امر به دلیل ویژگی قدرتمند شبکه CNN در استخراج ویژگی‌ها می‌باشد. همچنین شبکه عصبی عمیق BiLSTM نیز به دلیل پردازش توالی‌ها در دو جهت سبب می‌گردد کارایی مدل افزایش یابد. شکل (۸)، میزان صحت شبکه دستیل‌برت+ BiLSTM در حالت آموزش و ارزیابی نشان داده شده است.



شکل ۸- معیار دقت شبکه عصبی عمیق ترکیبی BiLSTM+ دستیل‌برت

در نمودار شکل (۹)، نرخ از دست رفتگی داده شبکه دستیل‌برت+BiLSTM در حالت آموزش و ارزیابی در ده دور نشان داده شده است.



شکل ۹- نرخ از دست رفتگی داده مدل دستیل‌برت+BiLSTM

همچنین، با اجرای مدل BiLSTM+دستیل‌برت برروی مجموع دادگان Res2015 و Res2016 نتایج جدول (۱۱) بدست آمده است. این نتایج در مقایسه با پژوهش‌های مشابه امیدوار کننده می‌باشد. ما مدل پیشنهادی را با پژوهش‌های زیر مقایسه کرده‌ایم.

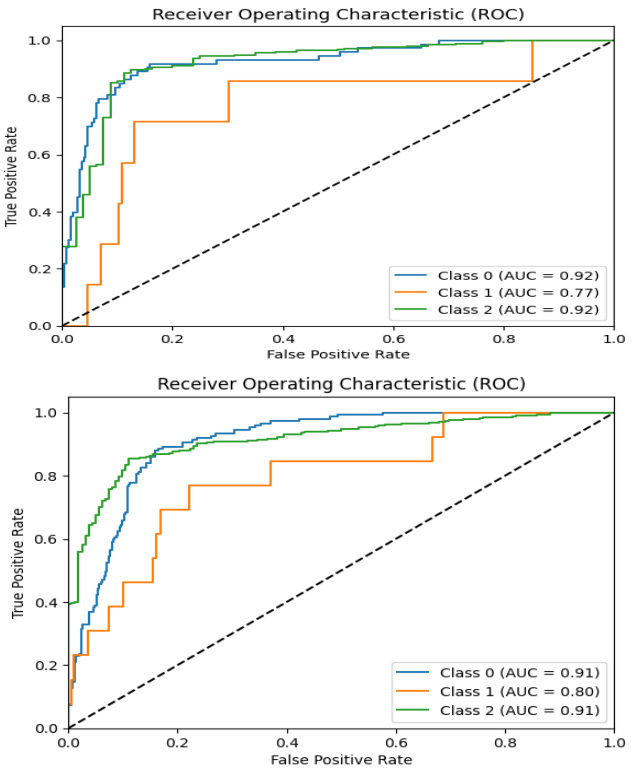
- CPA-SA [۴۰]: این مدل از دو تابع وزنی موقعیت متن، برای تنظیم اهمیت کلمات متنی بر اساس موقعیت آنها نسبت به کلمات جنبه در جمله‌ها استفاده می‌کند. این رویکرد به کاهش تأثیر عدم تعادل کلمات بر قضاوت احساسات ناشی از تعداد کلمات متفاوت در دو طرف کلمات جنبه کمک می‌کند. با استفاده از لایه‌های GRU دوطرفه (BiGRU) در هر دو سطح تک و چند جمله‌ای، این مدل تأثیر متنی هر جمله در سند را بر قطبیت احساسات جملات فردی به تصویر می‌کشد. علاوه بر این، این مقاله با تجزیه و تحلیل ویژگی‌های توزیع نمونه‌های چالشی و معرفی یک تابع ضرر جدید، راه‌حلی برای مسئله عدم تعادل قطبیت‌ها در تحلیل احساسات ارائه می‌دهد.
- TAS-BERT [۵۲]: این روش از مدل زبانی از پیش آموزش‌دیده برتا برای نشان داده وابستگی بین اهداف، جنبه‌ها و احساسات استفاده می‌کند.
- ASGCN [۴۰]: از شبکه پیچشی گراف محور (GCN) برای گرفتن اطلاعات معنایی و ایجاد ارتباط بین عناصر استفاده می‌کند.
- THA+STN [۵۲]: THA + STN، مدلی عصبی برای تشخیص احساسات هدفمند (TD) است. این مدل دارای دو لایه توجه خطی و لایه‌های کاملاً متصل (FC) است.
- MSRL-Net [۵۳]: نویسندگان مدلی بر پایه برت به نام شبکه یادگیری تقویت‌شده معنایی چندسطحی (MSRL-Net) برای تحلیل احساسات جنبه‌محور پیشنهاد داده‌اند. در MSRL-Net، برای انجام وظیفه تحلیل احساسات جنبه‌محور از روابط وابستگی کلمه و روابط کلمه–جمله برای تقویت نمایش معنایی سطح کلمه در تطبیق معنایی جمله استفاده شده است.

نتایج این مقایسه در جدول (۱۱) ذکر شده است.

جدول (۱۱): مقایسه مدل BiLSTM+دستیل‌برت و مقالات پیشین

مدلها	Res2015		Res2016	
	دقت	F1	دقت	F1
CPA-SA	۷۹/۵۷	۶۰/۲۶	۸۸/۸	۷۱/۴۷
TAS-BERT	-	۷۰/۴۲	-	۷۰/۶۶
BERT-pair-NLI-B	-	۷۰/۷۸	-	۸۰/۲۵
ASGCN	۷۹/۳۴	۶۰/۷۸	۸۸/۶۹	۶۶/۶۴
THA + STN	–	۷۱/۴۶	-	۷۳/۶۱
MSRL-Net	–	۹۲/۶۸	–	۸۶/۵۸
+BiLSTM دستیل برت	**۸۷/۶۱**	**۹۳/۱**	۸۵/۱۴	**۸۶/۸۶**

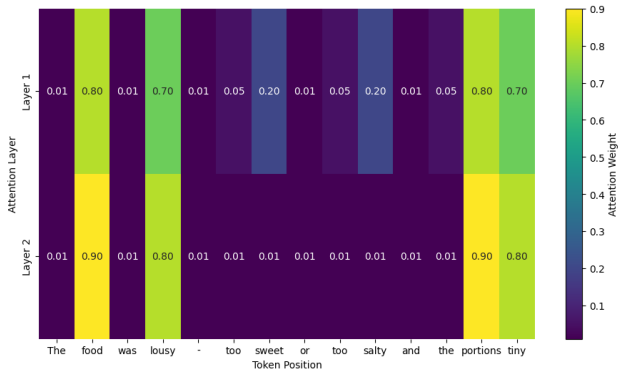
جدول (۱۱) نشان داده‌شده که مدل پیشنهادی در هر دو مجموعه داده Res2015 و Res2016 عملکرد پایداری دارد و F1 و دقت بسیار نزدیک است. در حالی که در مدل‌های دیگر، این دو معیار فاصله نسبتا زیادی با یکدیگر دارند که نشان می‌دهد این مدل‌ها دارای ایی در معیارهای دقت یا یادآوری هستند. علاوه بر این، عملکرد مدل در هر دو مجموعه داده نزدیک به یکدیگر و در سطح بهینه است. در مقابل، عملکرد مدل‌های دیگر مانند ASGCN، در یک مجموعه داده مطلوب و در مجموعه داده دیگری کاهش یافته است. در مجموع مدل پیشنهای به صورت میانگین دقت و F1 را بر روی مجموعه داده Res2015 در حدود ۴ درصد و F1 را بر روی مجموعه داده Res2016 در حدود ۱ درصد افزایش داده است. در شکل (۱۰) معیار ROC-AUC را برای درک بهتر کارایی مدل BiLSTM+دستیل‌برت نشان می‌دهد.



شکل ۱۰- معیار ROC-AUC شبکه عصبی عمیق ترکیبی BiLSTM+ دستیل‌برت (شکل اول Res2015 و دوم Res2016)

در شکل (۱۰)، کلاس «۰» برابر با قطب مثبت، کلاس «۱» برابر با قطبیت خنثی و کلاس «۲» برابر با قطبیت منفی است. در مجموعه داده های Res2015 و ۲۰۱۶، کلاس خنثی کمترین تعداد توییت را دارد. شکل (۱۰) نشان می‌دهد که مدل هر سه قطب مثبت، منفی و خنثی را برای جنبه های مختلف متن تشخیص داده است. همچنین، نرخ ROC-AUC برای مجموعه داده Res2015 برابر با ۸۵/۴ و برای Res2016 برابر با ۸۷/۳ است.

ما همچنین از نمودار نقشه حرارتی در شکل (۱۱) برای نشان دادن تأثیر لایه توجه بر روی متنی از دادگان SemEval استفاده کردیم. هر سلول وزن لایه‌های توجه مدل را نشان می‌دهد که به نشانه‌ها (کلمات) خاص طول پردازش متن اختصاص داده شده است. این وزن‌ها نشان می‌دهند که مدل کدام بخش از متن را هنگام تعیین احساسات بر اساس جنبه‌های خاص مهم‌تر می‌داند.



شکل ۱۱- نقشه حرارتی وزن‌های لایه توجه

همانطور که در شکل (۱۱) مشاهده می‌شود، نقشه حرارتی به طور موثر نشان می‌دهد که چگونه مکانیسم لایه توجه در مدل جمله "غذا کثیف بود - خیلی شیرین یا خیلی شور و قسمت‌های کوچک" را پردازش می‌کند. ردیف بالا که نمایانگر لایه ۱ است، تمرکز اولیه مدل را بر روی کلمات کلیدی مانند "food" و "lousy" نشان می‌دهد و شناسایی جنبه‌های حیاتی و

احساسات مرتبط با آنها را برجسته می‌کند. این لایه همچنین توجه به کلمات "portions" و "tiny" را نشان می‌دهد که نشان‌دهنده توانایی مدل برای گرفتن احساسات مربوط به جنبه‌های مختلف در یک جمله است. در لایه ۲، که در ردیف پایین نشان داده شده است، افزایش قابل توجهی در وزن توجه برای این کلمات وجود دارد که نشان دهنده تجزیه و تحلیل عمیق و تأیید ارتباط آنها است. این افزایش تمرکز از لایه ۱ به لایه ۲ نشان می‌دهد که این مدل نشانه‌های لازم را شناسایی کرده و تفسیر خود را با در دسترس قرار گرفتن اطلاعات زمینه‌ای بیشتر از طریق شبکه، اصلاح می‌کند.

بنابراین می‌توان به سوالات پژوهشی ذکر شده در مقدمه اینگونه پاسخ داد:

۱- چگونه می‌توان دادگان میان دامنه‌ای مناسبی در دامنه مبدأ انتخاب کرد که انتقال ویژگی‌های آنها به دامنه هدف موجب بروز پدیده انتقالی منفی نشود؟ با توجه به افزایش ابزارهای مانند API ها و رصد گسترده داده‌های موجود در وب، اگر در دامنه‌ای خاص داده کمیاب باشد می‌توان از دانش موجود در دامنه‌های مشابه استفاده کرد. با این حال، اگر دامنه‌ها مشابه نبوده یا شباهت پائینی داشته باشند، استفاده از دادگان میان‌دامنه‌ای ممکن است موجب انتقال منفی گردد. یکی از روش‌های موثری که در این پژوهش پیشنهاد شده است استفاده از ضریب شباهت پیروسون است. نتایج جدول (۶) و نمودار شکل (۳) نشان می‌دهد اگر امتیاز مرتبط با شباهت کسینوسی داده مبدا و هدف کم باشد، انتقال دانش موجب بروز پدیده انتقال منفی می‌گردد.

۲- چگونه می‌توان با ترکیب شبکه‌های یادگیری عمیق، لایه توجه و آموزش آنها با دادگان میان‌دامنه‌ای مدلی مطلوب با کمترین افت کارایی برای وظیفه تحلیل احساسات جنبه‌محور ارائه داد؟ شبکه‌های یادگیری عمیق هریک برای وظیفه‌ای خاص طراحی شده است. برای مثال CNN برای پردازش تصویر به وجود آمد ولی به دلیل ویژگی قدرتمند آن در استخراج ویژگی در سایر وظایف مانند پردازش زبان طبیعی نیز بکار گرفته شد [۵۱]. از همین رو، می‌توان بر ترکیب قابلیت‌های سودمند شبکه‌های عصبی عمیق سبب پایداری بیشتر مدل یادگیری گردید. برای نمونه شبکه‌های عصبی عمیق بازگشتی مانند BiLSTM برای پردازش توالی‌ها بسیار مناسب هستند. با این حال، این شبکه‌ها نیازمند دادگان بسیار غنی جهت آموزش می‌باشند. از طرف دیگر، در دامنه‌های خاص مانند رمزارزها دادگان غنی بسیار نادر می‌باشد. همچنین، این شبکه‌ها (مانند BiLSTM) قدرت بالایی در استخراج ویژگی‌ها ندارند. لذا، همانگونه که در معماری مدل پیشنهادی شکل (۵) نشان داده شده می‌توان از شبکه‌های پیش‌آموزش دیده مانند دستیل‌برت به عنوان ستوان فقرات شبکه، جهت استخراج ویژگی و انتقال دانش استفاده کرد. با دانش استخراج شده و تنظیم شبکه دستیل‌برت می‌توان این دانش را به BiLSTM انتقال داد و آنرا جهت انجام وظیفه در دامنه هدف آموزش داد. کارایی این روش در جدول (۹)، اثبات شده است.

۳- با توجه به اینکه استفاده از مدل‌های یادگیری عمیق پایه و پیش‌آموزش‌دیده برای استخراج ویژگی و انتقال دانش از دامنه مبدأ به هدف مؤثر می‌باشد. کدامیک از این مدل‌ها برای تحلیل احساسات رمزارزها اثربخشی بیشتری دارد؟ همانگونه که ذکر شد شبکه‌های عصبی عمیق، مدل‌های حریص به داده هستند. این شبکه‌ها برای کارایی مطلوب نیازمند دادگان غنی برچسب دار می‌باشند. در تحلیل احساسات رمزارزها دادگان غنی برچسب‌دار مخصوصاً همراه با جنبه‌های متن که مرتبط با انواع رمزارزها باشد، کمیاب است. از این رو، آموزش شبکه‌های عصبی پایه در دامنه رمزارزها امری چالش بر انگیز می‌باشد. لذا،

می‌توان از شبکه‌های عصبی پیش‌آموزش دیده مانند انواع برت که توسط مجموعه برزگی از دادگان جهت وظایف مرتبط با پردازش زبان طبیعی تنظیم شده‌اند، استفاده کرد. بنابراین، همانگونه که در جدول (۹)، نشان داده شده ترکیب شبکه‌های عصبی عمیق پیش‌آموزش‌دیده ساده و ترکیبی منجر به دست‌یابی به نتایج مطلوب‌تری نسب به شبکه‌های عصبی عمیق پایه شده است.

٤- دادگان نامتوازن سبب جهت‌گیری مدل به سمت قطبیتی خاص می‌شود، چگونه می‌توان با این مشکل برخورد کرد؟ برای رفع مشکل دادگان نامتوازن در این پژوهش از تکنیک SMOTE استفاده شده است. SMOTE تکنیکی آماری است که به طور مصنوعی نمونه‌های جدیدی از کلاس اقلیت را در فضای ویژگی ایجاد می‌کند تا مجموعه داده را متعادل کند. این تکنیک سبب می‌شود توازن قطبیت‌ها در کل دادگان حفظ شده و مدل در فرایند یادگیری به سمت قطبیتی خاص جهت‌گیری نداشته باشد.

در بخش نتایج مقاله، مدل پیشنهادی با چندین روش مقایسه شده و بهبودهای چشمگیری در دقت و عملکرد کلی مشاهده شده است. نتایج نشان می‌دهد مدل پیشنهادی توانسته به صورت میانگین دقت و F1 را بر روی مجموعه‌داده بنچ‌مارک SemEval در حدود ۲ درصد افزایش داده و میانگین نرخ ROC-AUC به ۸۶,۳۵ درصد رسیده است. این نتایج به دلیل استفاده از یادگیری انتقالی شات صفر، بهینه‌سازی داده‌های ورودی و ترکیب مدل‌های پیش‌آموزش‌دیده با شبکه‌های عصبی عمیق حاصل شده است که برخی از بهبودها در زیر ذکر شده است.

- **عملکرد بهبود یافته در مقایسه با مدل‌های سنتی:** نتایج نشان می‌دهند که مدل پیشنهادی به‌طور قابل‌توجهی عملکرد بهتری نسبت به مدل‌های سنتی مانند LSTM و GRU دارد. این بهبود به دلیل استفاده از داده‌های متنوع و مرتبط با همبستگی بالا با موضوع مورد بحث است که به مدل کمک می‌کند تا درک عمیق‌تری از داده‌های ورودی داشته باشد. تفاوت اصلی این است که مدل‌های سنتی به داده‌های برچسب‌گذاری شده بیشتر و تنظیمات پیچیده‌تری نیاز دارند، در حالیکه مدل پیشنهادی با استفاده از یادگیری انتقالی و داده‌های کم‌برچسب‌گذاری شده یا بدون برچسب‌گذاری به نتایج مطلوب‌تری دست یافته است.
- **بهبود دقت در تحلیل احساسات جنبه‌محور:** یکی از مهم‌ترین نتایج به‌دست‌آمده، بهبود دقت در تحلیل احساسات جنبه‌محور است. مدل پیشنهادی با استفاده از مکانیسم‌های توجه و ترکیب داده‌های متنوع، قادر به شناسایی بهتر روابط بین جنبه‌ها و احساسات در متن‌ها شده است. این بهبود به دلیل استفاده از داده‌های همبسته و مرتبط با موضوع و همچنین استفاده از شبکه‌های عمیق پیش‌آموزش‌دیده بوده که توانایی مدل را در تحلیل دقیق‌تر افزایش داده است.
- **کاهش پدیده انتقال منفی:** یکی از مشکلات رایج در یادگیری انتقالی، انتقال منفی است که می‌تواند دقت مدل را کاهش دهد. در این تحقیق، با استفاده از داده‌هایی با همبستگی بالا و تنظیم ابرپارامترها، این پدیده به‌طور قابل‌توجهی کاهش یافته است. این انتخاب دقیق داده‌ها باعث شده است که مدل نه‌تنها دقت بالایی در دامنه‌های مختلف داشته باشد، بلکه توانایی تعمیم‌پذیری بیشتری را نیز نشان دهد.

- **مقایسه با روش‌های پیشرفته‌تر:** در مقایسه با روش‌های پیشرفته مانند GCN و مدل‌های ترکیبی، مدل پیشنهادی توانسته است با ساده‌سازی فرآیند آموزش، استفاده از داده‌های مرتبط، تنظمی ابرپارامترها و نرخ یادگیری وفق‌پذیر، عملکردی مطلوب ارائه دهد. این تفاوت به دلیل استفاده از یادگیری انتقالی و بهینه‌سازی‌های انجام‌شده در مدل است که باعث شده است تا مدل نه‌تنها دقت بیشتری داشته باشد، بلکه کارایی بالاتری از نظر معیار F1 را نیز نشان دهد.

۵. نتیجه‌گیری و کارهای آینده

رویکردهای اولیه پردازش زبان طبیعی با استفاده از CNN، شبکه عصبی بازگشتی و BiLSTM دارای محدودیت‌هایی برای درک بافت عمیق‌تر جملات هستند. از این رو، شبکه پیش‌آموزش‌دیده برت درک بهتری از محتوای جملات نسبت به رویکردهای پایه فراهم می‌آورد. زیرا، در لایه‌های رمزگذار شبکه برت همه ورودی‌ها (کل جمله) به طور همزمان پردازش می‌شوند. بنابراین، هنگام ساختن زمینه برای یک کلمه، برت ورودی‌های قبل از آن و همچنین ورودی‌های بعدی را در نظر می‌گیرد. همچنین، نتایج نشان می‌دهد، استفاده از روش ضریب شباهت پیرسون جهت انتخاب مجموعه دادگان دامنه مبدا انتقال منفی دانش به دامنه هدف را به صورت چشم‌گیری کاهش می‌دهد. علاوه بر این، نرخ یادگیری وفق‌پذیر مورد استفاده در این پژوهش و انتخاب ابرپارامترها به روش گریدی سبب گردیده شبکه‌های عصبی عمیق در تعداد تکرار پائین‌تری به همگرایی برسند که بسیار در کاربردهای تجاری، حیاتی می‌باشد. همچنین در این پژوهش نشان داده شد، استفاده از مدل‌های تنظیم‌شده مبنی بر برت با ترکیب شبکه‌های عصبی عمیق BiLSTM می‌تواند کارایی مطلوبی در تحلیل احساسات دامنه‌های نوظهور مانند رمزارزها داشته باشد. از سوی دیگر، این پژوهش با استفاده از یادگیری انتقالی با شات صفر و صرفاً استفاده از دادگان میان‌دامنه‌ای مشابه نتایج امیدوار کننده‌ای را برای تحلیل احساسات جنبه‌محور در دامنه‌های خاص بدست آورده است. در تحقیقات آینده، می‌توان به دنبال پاسخ به این سوال بود که چه میزان از انتقال دانش سبب بهبود مدل در دامنه هدف می‌شود. به عبارت دیگر، تا چه میزان افزایش مجموعه دادگان در دامنه مبدا کارایی مدل در دامنه هدف را بهبود می‌بخشد. همچنین می‌توان بررسی کرد آیا استفاده از لایه‌های توجه در شبکه‌های عصبی عمیق ترکیبی پایه مانند CNN+BiLSTM یا اشتراک‌گذاری وزن‌ها در شبکه‌های کانولوشنی می‌تواند سبب شود این مدل‌ها به دقت‌های مشابهی با شبکه‌های عصبی پیش‌آموزش دیده دست یابند. علاوه بر این، می‌توان لایه‌های منجمد شده شبکه‌های پیش‌آموزش دیده را به عنوان یک ابرپارامتر برای بهینه‌سازی مورد بررسی قرارداد. همچنین می‌توان آزمایش کرد ترکیب ویژگی‌های استخراج شده از مدل‌های تعبیه کلمات مانند Word2Vec با وزن‌های استخراج شده از شبکه‌های پیش‌آموزش‌دیده چه تاثیری بر کارایی مدل‌های عمیق دامنه هدف دارد.

مراجع

[۱] J. Chai and A. Li, “Deep learning in natural language processing: A state-of-the-art survey,” in *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2019. DOI: 10.1109/ICMLC48188.2019.8949185.

[۲] F. Zare Mehrjardi, M. Yazdian-Dehkordi, and A. Latif, "Evaluating classical machine learning and deep-learning methods in sentiment analysis of Persian telegram message", *Soft Computing* Journal, vol. 11, no. 1, pp. 88-105, 2022. DOI: https://doi.org/10.22052/scj.2023.246553.1077.

[۳] O. Khalaf Beigi, S. A. Bashiri Mosavi, and S. Gharloghi, "Applying Character-Level Neural Network-Based Sentiment Analysis Model on Persian Comments of the Social Media-Online Store Platforms", *Soft Computing Journal*, vol. 11, no. 2, pp. 118-133, 2023 DOI: https://doi.org/10.22052/scj.2023.248311.1094.

[٤] F. Pourgholamali, M. Kahani, and E. Asgarian, "Exploiting Big Data Technology for Opinion Mining ", *Soft Computing Journal*," vol. 9, no.

617–663, 2019. DOI: https://doi.org/10.1007/s10115-018-1236-4.

[١٦] Z. Nanli, Z. Ping, L. Weiguo, and C. Meng, "Sentiment analysis: A literature review," in *2012 International Symposium on Management of Technology (ISMOT)*, 2012 ,pp. 572-576: IEEE. DOI: DOI: 10.1109/ISMOT.2012.6679538.

[١٧] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-based sentiment analysis: A survey of deep learning methods," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 6, pp. 1358–1375, 2020. DOI: 10.1109/TCSS.2020.3033302.

[١٨] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 249–270, 2022. DOI: 10.1109/TKDE.2020.2981333.

[١٩] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019. DOI: 10.1109/ACCESS.2019.2912200.

[٢٠] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019. DOI: https://doi.org/10.1162/neco_a_01199.

[٢١] N. Agarwal, A. Sondhi, K. Chopra, and G. Singh, "Transfer learning: Survey and classification," in *Smart Innovations in Communication and Computational Sciences*, Singapore: Springer Singapore, 2021, pp. 145–155. DOI: https://doi.org/10.1007/978-981-15-5345-5_13.

[٢٢] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional Transformers for language understanding," *arXiv [cs.CL]*, 2018. DOI: https://doi.org/10.48550/arXiv.1810.04805.

[٢٣] I. Staliūnaitė and I. Iacobacci, "Compositional and lexical semantics in RoBERTa, BERT and DistilBERT: A case study on CoQA," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020. DOI: https://doi.org/10.48550/arXiv.2009.08257.

[٢٤] A. F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of Bert, Roberta, distilbert, and xlnet for text-based emotion recognition," in *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2020. DOI: 10.1109/ICCWAMTIP51612.2020.9317379.

[٢٥] M. Suhasini and B. Srinivasu, "Emotion detection framework for twitter data using supervised classifiers," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Nature Singapore, 2020, pp. 565–576. DOI: https://doi.org/10.1007/978-981-15-1097-7_47.

1, pp. 26-39, 2021. DOI: https://doi.org/10.22052/scj.2021.111450.

[٥] K. W. Church, Z. Chen, and Y. Ma, "Emerging trends: A gentle introduction to fine-tuning," *Nat. Lang. Eng.*, vol. 27, no. 6, pp. 763–778, 2021. DOI: https://doi.org/10.1017/S1351324921000322.

[٦] K. Jahanbin and M. A. Z. Chahooki, "Database of Twitter influencers in cryptocurrency (2021-2023) with sentiments," *Research Square*, 2023. https://doi.org/10.21203/rs.3.rs-3192598/v1.

[٧] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, 2024. DOI: https://doi.org/10.1007/s10994-022-06296-4.

[٨] G. Lu, Y. Liu, J. Wang, and H. Wu, "CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels," *Inf. Process. Manag.*, vol. 60, no. 3, p. 103320, 2023. DOI: https://doi.org/10.1016/j.ipm.2023.103320.

[٩] H. N. Serere, B. Resch, and C. R. Havas, "Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection," *PLoS One*, vol. 18, no. 3, p. e0282942, 2023. DOI: https://doi.org/10.1371/journal.pone.0282942

[١٠] M. Isnan, G. N. Elwirehardja, and B. Pardamean, "Sentiment analysis for TikTok review using VADER sentiment and SVM model," *Procedia Computer Science*, vol. 227, pp. 168-175, 2023. DOI: https://doi.org/10.1371/journal.pone.0282942.

[١١] M. Nair, L. A. Abd-Elmegid, and M. I. Marie, "Sentiment analysis model for cryptocurrency tweets using different deep learning techniques," *J. Intell. Syst.*, vol. 33, no. 1, 2024. DOI: https://doi.org/10.1515/jisys-2023-0085.

[١٢] T. Pano and R. Kashef, "A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19," *Big Data Cogn. Comput.*, vol. 4, no. 4, p. 33, 2020. DOI: https://doi.org/10.3390/bdcc4040033.

[١٣] A. K. Jha and N. K. Verma, "Social media sustainability communication: An analysis of firm behaviour and stakeholder responses," *Inf. Syst. Front.*, 2022. DOI: https://doi.org/10.1007/s10796-022-10257-6

[١٤] N. K. Singh, D. S. Tomar, and A. K. Sangaiah, "Sentiment analysis: a review and comparative analysis over social media," J. Ambient Intell. Humaniz. Comput., vol. 11, no. 1, pp. 97–117, 2020. DOI: https://doi.org/10.1007/s12652-018-0862-8.

[١٥] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, no. 2, pp.

convolutional networks for aspect-level sentiment classification," *Knowl. Based Syst.*, vol. 193, no. 105443, p. 105443, 2020. DOI: https://doi.org/10.1016/j.knosys.2019.105443.

[٣٦] A. Žunić, P. Corcoran, and I. Spasić, "Aspect-based sentiment analysis with graph convolution over syntactic dependencies," *Artif. Intell. Med.*, vol. 119, no. 102138, p. 102138, 2021. DOI: https://doi.org/10.1016/j.artmed.2021.102138.

[٣٧] Y. Lv, F. Wei, L. Cao, S. Peng, J. Niu, S. Yu, and C. Wang, "Aspect-level sentiment analysis using context and aspect memory network," *Neurocomputing*, vol. 428, pp. 195–205, 2021. DOI:https://doi.org/10.1016/j.neucom.2020.11.049.

[٣٨] N. Habbat, H. Anoun, and L. Hassouni, "Combination of GRU and CNN deep learning models for sentiment analysis on French customer reviews using XLNet model," *IEEE Engineering Management Review*, vol. 51, no. 1, pp. 41-51, 2022. DOI: 10.1109/EMR.2022.3208818.

[٣٩] M. M. Kabir, Z. A. Othman, M. R. Yaakub, and S. Tiun, "Hybrid syntax dependency with lexicon and logistic regression for aspect-based sentiment analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, 2023. DOI: http://dx.doi.org/10.14569/IJACSA.2023.0141085.

[٤٠] B. Huang, R. Guo, Y. Zhu, Z. Fang, G. Zeng, J. Liu, Y. Wang, H. Fujita, and Z. Shi, "Aspect-level sentiment analysis with aspect-specific context position information," *Knowledge-Based Systems*, vol. 243, p. 108473, 2022. DOI: https://doi.org/10.1016/j.knosys.2022.108473.

[٤١] X. Zhu, L. Zhu, J. Guo, S. Liang, and S. Dietze, "GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification," Expert Syst. Appl., vol. 186, no. 115712, p. 115712, 2021. DOI: https://doi.org/10.1016/j.eswa.2021.115712.

[٤٢] S. Chumakov, A. Kovantsev, and A. Surikov, "Generative approach to aspect based sentiment analysis with GPT language models," Procedia Comput. Sci., vol. 229, pp. 284–293, 2023. DOI: https://doi.org/10.1016/j.procs.2023.12.030.

[٤٣] P. Liashchynskyi, "Grid Search, Random Search, Genetic Algorithm: A big comparison for NAS," *arXiv [cs.LG]*, 2019. DOI: https://doi.org/10.48550/arXiv.1912.06059.

[٤٤] K. Nakamura, B. Derbel, K.-J. Won, and B.-W. Hong, "Learning-rate annealing methods for deep neural networks," *Electronics (Basel)*, vol. 10, no. 16, p. 2029, 2021. DOI: https://doi.org/10.3390/electronics10162029.

[٤٥] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artif. Intell.*

[٢٦] Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," SMU Data Science Review, vol. 1, no. 3, p. 1, 2018. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/1.

[٢٧] S. Biswas, M. Pawar, S. Badole, N. Galande, and S. Rathod, "Cryptocurrency price prediction using neural networks and deep learning," in *2021 7th international conference on advanced computing and communication systems (ICACCS)*, 2021, vol. 1 ,pp. 408-413: IEEE. DOI: 10.1109/ICACCS51430.2021.9441872.

[٢٨] X. Huang *et al.*, "Lstm based sentiment analysis for cryptocurrency prediction," in *Database Systems for Advanced Applications: 26th International Conference, DASFAA 2021, , 2021, Proceedings*, pp. 617-621. DOI: https://doi.org/10.1007/978-3-030-73200-4_47.

[٢٩] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, 2014, vol. 8, no. 1, pp. 216-225. DOI: https://doi.org/10.1609/icwsm.v8i1.14550

[٣٠] S. Oikonomopoulos, K. Tzafilkou, D. Karapiperis, and V. Verykios, "Cryptocurrency price prediction using social media sentiment analysis," in *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 2022, pp. 1-8.IEEE. DOI: https:// 10.1109/IISA56318.2022.9904351.

[٣١] G. C. Habek, M. A. Toçoğlu, and A. Onan, "Bi-directional CNN-RNN architecture with group-wise enhancement and attention mechanisms for cryptocurrency sentiment analysis," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022. DOI: https://doi.org/10.1080/08839514.2022.2145641

[٣٢] I. Moudhich and A. Fennan, "Graph embedding approach to analyze sentiments on cryptocurrency," Int. J. Electr. Comput. Eng. (IJECE), vol. 14, no. 1, p. 690, 2024. DOI: http://doi.org/10.11591/ijece.v14i1.pp690-697.

[٣٣] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022. DOI: https:// 10.1109/ACCESS.2022.3165621

[٣٤] J. Davchev, K. Mishev, I. Vodenska, L. Chitkushev, and D. Trajanov, "Bitcoin price prediction using transfer learning on financial micro-blogs," in *The 16th Annual International Conference on Computer Science and Education in Computer Science*, 2020. Available at: https://www.ceeol.com/search/article-detail?id=978526

[٣٥] P. Zhao, L. Hou, and O. Wu, "Modeling sentiment dependencies with graph

Rev., vol. 55, no. 7, pp. 5543–5574, 2022. DOI: https://doi.org/10.1007/s10462-021-10134-9.

[٤٦] A. J. Amadeo, J. G. Siento, T. A. Eikwine, Diana, and I. H. Parmonangan, "Temporal Fusion Transformer for Multi Horizon Bitcoin Price Forecasting," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, 2023. DOI: https://doi.org/10.1109/ITIS59651.2023.10420330

[٤٧] A. Y. Nouira, M. Bouchakwa, and M. Amara, "Role of social networks and machine learning techniques in cryptocurrency price prediction: a survey," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, 2024. DOI: https://doi.org/10.1007/s13278-024-01316-8

[٤٨] G. Serafini *et al.*, "Sentiment-driven price prediction of the bitcoin based on statistical and deep learning approaches," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020. DOI: https://doi.org/10.1109/IJCNN48605.2020.9206704

[٤٩] N. Elhadad, S. Pradhan, S. Gorman, S. Manandhar, W. Chapman, and G. Savova, "SemEval-2015 Task 14: Analysis of Clinical Text," *in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015. DOI: https://doi.org/10.18653/v1/S15-2051

[٥٠] M. Pontiki et al., "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," *in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016. DOI: https://aclanthology.org/S16-1002

[٥١] 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: https://doi.org/10.3390/make1030048.

[٥٢] H. Wan, Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," *Proc. Conf. AAAI Artif. Intell.*, vol. 34, no. 05, pp. 9122–9129, 2020. DOI: https://doi.org/10.1609/aaai.v34i05.6447.

[٥٣] Z. Hu, Z. Wang, Y. Wang, and A.-H. Tan, "MSRL-Net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis," *Expert Syst. Appl.*, vol. 217, no. 119492, p. 119492, 2023. DOI: https://doi.org/10.1016/j.eswa.2022.119492.