بهبود نظرکاوی فارسی مبتنی بر قطبیت و متوازن‌سازی کلمات مهم مثبت و منفی (مطالعه موردی: نظرات دیجی‌کالا برای موبایل)

مهدیه واحدی پور، محبوبه شمسی و عبدالرضا رسولی کناری
دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی قم، قم – ایران

چکیده

بسیاری از شبکه‌های اجتماعی و سایت‌ها به مردم اجازه می‌دهند تا احساسات و نظرات خود را در مورد محصولات و خدمات مختلف به اشتراک بگذارند. در این مقاله روشی جدید مبتنی بر قطبیت نظرات مثبت و منفی فارسی درباره محصولات تلفن همراه از سایت دیجی‌کالا و داده‌های سنتی‌پرس ارائه شده است. نتیجه اجرا با الگوریتم‌های بیز ساده، ماشین بردار پشتیبان، کاهش گرادیان تصادفی، رگرسیون لجستیک، جنگل تصادفی و یادگیری عمیق مانند شبکه عصبی کانولوشن و حافظه کوتاه‌مدت متوالی بر اساس پارامترهایی مانند صحت، بازیابی، معیار فیشر و دقت، موردتوجه قرار گرفته شده است. روش پیشنهادی روی داده‌های دیجی‌کالا، با الگوریتم‌های بیز ساده بین ۱۰ تا ۳۴ درصد و ماشین بردار پشتیبان بین ۵ تا ۲۴ درصد و کاهش گرادیان تصادفی بین ۷ تا ۳۸ درصد و رگرسیون لجستیک بین ۵ تا ۳۸ درصد و جنگل تصادفی بین ۴ تا ۲۲ درصد و روش شبکه عصبی کانولوشن به میزان ۴ درصد افزایش دقت را به همراه داشته است. هم‌چنین در داده‌های سنتی‌پرس با الگوریتم‌های بیز ساده بین ۱۲ تا ۴۶ درصد و ماشین بردار پشتیبان بین ۵ تا ۴۶ درصد و کاهش گرادیان تصادفی بین ۵ تا ۳۵ درصد و رگرسیون لجستیک بین ۶ تا ۴۶ درصد و جنگل تصادفی بین ۴ تا ۴۶ درصد دقت نسبت به قبل از اعمال روش پیشنهادی به‌دست آمده است.

واژگان کلیدی: تحلیل احساسات، نظرکاوی، یادگیری ماشین، یادگیری عمیق، قطبیت

Improving Persian Opinion Mining based on Polarity and Balancing Positive and Negative Keywords (case study: Digikala reviews for mobile)

Mahdieh Vahedipoor[1], Mahboubeh Shamsi and Abdolreza Rasouli Kenari
Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran

Abstract:
In recent years, the massive growth of generated content by the users in social networks and online marketing sites, allows people to share their feelings and opinions in a variety of opinions about different products and services. Sentiment analysis is an important factor for better decision-making that is done using natural language processing (NLP), computational methods, and text analysis to extract the polarity of unstructured documents. The complexity of human languages and sentiment analysis have created a challenging research context in computer science and computational linguistics. Many researchers used supervised machine learning algorithms such as Naïve Bayes (NB), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Logistic Regression (LR) Random Forest (RF), and deep learning algorithms such as Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM). Some researchers have used Dictionary-based methods. Despite the existence of effective techniques in text mining, there are still unresolved challenges. Note that user comments are

* نویسندهٔ عهده‌دار مکاتبات
* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۱۰/۱۱ • تاریخ پذیرش: ۱۴۰۲/۹/۲۰ • تاریخ انتشار: ۱۴۰۲/۱۲/۲۹ • نوع مطالعه: پژوهشی

unstructured texts; Therefore, in order to structure the textual inputs, parsing is usually done along with adding some features, linguistic interpretations and removing additional items, and inserting the next terms in the database, then extracting the patterns in the structured data and finally the outputs will evaluate and interpret. The imbalance of data with the difference in the number of samples in each class of a dataset is an important challenge in the learning phase. This phenomenon breaks the performance of the classifications because the machine does not learn the features of the unpopulated classes well. In this paper, words are weighted based on the prescribed dictionary to influence the most important words on the result of the opinion mining by giving higher weight. On the other hand, the combination of the adjacent words using n-gram methods will improve the outcome. The dictionaries are highly related to the domain of the application. Some words in an application are important but in mobile comments are not impressive. Another challenge is the unbalanced train data, in which the number of positive sentences is not equal to the number of negative sentences. In this paper, two ideas are applied to build an efficient opinion mining algorithm. First, we build a precise dictionary for mobile Persian comments, and the second idea is to balance the positive and negative comments in train data. In summary, the main achievements of the current research can be mentioned: creating a weighted comprehensive dictionary in the field of mobile phone opinions to increase the accuracy of opinion analysis, balancing positive and negative opinions to improve the accuracy of opinion analysis, and eliminating the negative effect of overfitting and providing a precise approach to Determining the polarity of users' opinions about mobile phones using machine learning and recurrent deep learning algorithms. This new method is presented on mobile phone products from the Digikala site and Senti-Pers data. The result is performed with Naive Bayesian, Support Vector Machine, Stochastic Gradient Descent, Logistic Regression, Random Forest, and deep learning methods such as Convolutional Neural Network and Long Short-Term Memory based on parameters such as Accuracy, Precision, Retrieval, and F-Measure. The proposed method increases accuracy on Digikala, with NB between 10% and 34% and SVM between 5% and 24%, SGD between 7% and 38%, LR between 5% to 38%, and RF between 4% Up to 22% and CNN by 4%. The results show an accuracy increment on Senti-Pers, with NB between 12% and 46% and SVM between 5% and 46%, SGD between 5% and 35%, LR between 6% to 46%, and RF between 4% Up to 46%.

Keywords: Sentiment Analysis, Opinion Mining, Machine Learning, Deep Learning, Polarity

۱- مقدمه

در حال حاضر شبکه‌ها و رسانه‌های اجتماعی نقش مهمی در ارائه اطلاعات در مورد هر محصول با استفاده از نظرات مختلف کاربران دارند. در علوم رایانه، تجزیه‌وتحلیل احساسات به‌عنوان نظرکاوی شناخته شده است و نظرات مردم و هم‌چنین احساسات آن‌ها نسبت به محصولات، سازمان‌ها و ویژگی‌های مربوط به آن‌ها را تحلیل می‌کند. محققان با روش‌های مختلف یادگیری ماشین و به‌منظور استخراج اطلاعات معنادار از احساسات مردم از این داده‌ها استفاده می‌کنند. تشخیص ساختار متن به سه شکل مختلف سند، جمله و کلمه امکان‌پذیر است. در سطح سند، اسناد به نظرات مثبت (موافق)، منفی (مخالف) و خنثی (بی‌طرف) دسته‌بندی می‌شود. در سطح جمله تعیین می‌کنند که آیا جمله نظر مثبت، منفی و یا خنثی را بیان می‌کند و در سطح ویژگی، برای تعیین نظر مثبت، منفی و یا خنثی به جزئی‌ترین شکل ممکن (کلمه) انجام می‌گیرد و نظرات را در مورد ویژگی‌های خاص موضوع بررسی می‌کنند [۱]. در این مقاله، داده‌های جمع‌آوری شده نظرات تلفن همراه از سایت دیجی‌کالا و داده‌های سنتی‌پرس از سایت پیکره‌گان تحلیل احساسات در سطح جمله موردتوجه قرار گرفته است.

به‌طور عمده دو نوع تکنیک یادگیری ماشین برای دسته‌بندی متون وجود دارد: تکنیک مبتنی بر یادگیری ماشین تحت نظارت و تکنیک مبتنی بر یادگیری ماشین بدون نظارت. در روش یادگیری نظارت‌شده، مجموعه‌داده‌ها توسط کارشناسان به‌صورت دستی یا با استفاده از فرهنگ لغت‌نامه برچسب‌گذاری می‌شوند [۵،۲۳]. بنابراین یادگیری مدل‌ها با بخشی از داده‌ها به‌عنوان داده‌های آموزشی صورت گرفته و با بخش دیگری از داده‌ها به‌عنوان داده‌های آزمایشی یک مدل ریاضی معقول را به دست می‌آورند [۲]. برخلاف یادگیری نظارت‌شده، در فرآیند یادگیری بدون نظارت هیچ اطلاعاتی در مورد برچسب وجود ندارد و نمی‌توان به‌راحتی پردازش کرد، زیرا آموزشی صورت نمی‌گیرد. الگوریتم‌های خوشه‌بندی با دسته‌بندی داده‌های مشابه جهت حل مشکل پردازش داده‌های بدون برچسب به کار می‌روند [۶].

از سوی دیگر الگوریتم‌های یادگیری عمیق، زیرمجموعه‌ای از الگوریتم‌های یادگیری ماشین هستند که هدف آن‌ها کشف چندین سطح از نمایش‌های توزیع‌شده از داده ورودی است. در سال‌های اخیر الگوریتم‌های یادگیری عمیق زیادی برای حل مسائل هوش مصنوعی سنتی ارائه شده‌اند. کاربردهای آن‌ها در زمینه‌های مختلف

بینایی ماشین همانند دسته‌بندی تصاویر، شناسایی اشیا، استخراج تصاویر، قطعه‌بندی معنایی و برآورد ژست انسان است [۳].

صرف‌نظر از الگوریتم مورد استفاده برای یادگیری، نظرات کاربران، متون بدون ساختار هستند؛ بنابراین، برای ساختاردهی به ورودی‌های متنی به طور معمول تجزیه، همراه با افزودن برخی ویژگی‌ها، تفاسیر زبانی و حذف موارد اضافی و درج موارد بعدی در پایگاه داده انجام می‌گیرد، سپس استخراج الگوهای درون داده‌های ساختاریافته و درنهایت ارزیابی و تفسیر خروجی‌ها انجام می‌شود. عدم تعادل داده‌ها با اختلاف در تعداد نمونه‌ها در هر کلاس از یک مجموعه‌داده چالش مهمی در مرحله یادگیری است. این پدیده عملکرد دسته‌بندی‌ها را خراب می‌کند، زیرا ماشین ویژگی‌های کلاس‌های با جمعیت کمتر را خوب نمی‌آموزد. استفاده از تکنیک‌های نمونه‌گیری ازجمله نمونه‌گیری بیش‌ازحد و نمونه‌گیری کمتر از حد عملکرد دسته‌بندی را بهبود می‌بخشند. برای مجموعه‌داده‌های کوچک، نمونه‌گیری بیش‌ازحد انتخاب شده است که مناسب‌ترین استراتژی است به دلیل این‌که مجموعه اولیه نمونه را افزایش می‌دهد و این کار، بر روی وظایف دودویی یا چند ستونی تمرکز دارد [۴]. در این مقاله از نمونه‌گیری بیش‌ازحد برای سازگاری داده‌ها استفاده شده است.

در این پژوهش، ابتدا کلمات توقف و سایر اطلاعات ناخواسته از نظرات کاربرانی که با استفاده از خزنده‌ها جمع‌آوری شده‌اند، حذف می‌گردد و سپس طبق فرهنگ‌نامه موردنظر ارزش نظرات، محاسبه شده و قطبیت جدیدی به نظرات داده می‌شود. به دلیل ناهمگن‌بودن داده‌های مثبت و منفی با روش نمونه‌گیری بیش‌ازحد، داده‌ها متعادل می‌شوند و از طریق فرآیند بردارسازی داده‌های متن ساختاریافته استخراجی از مرحله پیش‌پردازش به ماتریس اعداد تبدیل می‌شوند؛ سپس این ماتریس‌های عددی به الگوریتم‌های مختلف یادگیری ماشین تحت نظارت و یادگیری عمیق برای دسته‌بندی این نظرات داده شده و سپس پارامترهای مختلف مانند صحت، بازیابی، معیار فیشر و دقت برای ارزیابی عملکرد الگوریتم‌های یادگیری ماشین تحت نظارت و یادگیری عمیق جهت بهبود نظرکاوی نظرات فارسی مورد مقایسه قرار گرفته است. به‌طور خلاصه از دستاوردهای اصلی تحقیق حاضر می‌توان به موارد زیر اشاره کرد:

- ساخت یک فرهنگ‌نامه جامع وزن‌دار در حوزه نظرات تلفن همراه جهت افزایش دقت نظرکاوی
- متوازن‌سازی نظرات مثبت و منفی جهت بهبود دقت نظرکاوی و از بین بردن اثر منفی بیش‌برازش
- ارائه رویکرد دقیق جهت تعیین قطبیت نظرات کاربران در مورد تلفن همراه با استفاده از الگوریتم‌های یادگیری ماشین و یادگیری عمیق بازگشتی

ساختار مقاله به شرح زیر است: در بخش ۲، به بررسی ادبیات تحقیق پرداخته شده است. در بخش ۳، پیاده‌سازی رویکرد پیشنهادی توضیح داده شده است. در بخش ۴، ارزیابی عملکرد روش پیشنهادی انجام شده است و دقت رویکرد پیشنهادی با کارهای دیگران مقایسه شده است و در انتها در بخش ۵ مقاله نتیجه‌گیری و کارهای آینده پیشنهاد شده است.

۲- کارهای گذشته

تجزیه‌وتحلیل احساسات در حوزه وسیعی مانند بررسی فیلم، بررسی تدریس، بررسی محصولات، یادگیری الکترونیکی، بررسی هتل و بسیاری دیگر مورد مطالعه قرار گرفته است. بیشتر دانشمندان متمرکز به تجزیه‌وتحلیل داده‌های کمی هستند [۲۱-۲۵]. بااین‌حال، برخی از مطالعات انجام‌شده بر روی داده‌های کیفی با استفاده از تجزیه‌وتحلیل احساسات انجام‌شده است و هم‌چنین مطالعه خوبی توسط نویسندگان مختلف بر اساس دسته‌بندی احساسات در سطح سند انجام شده است که در زیر به برخی از آن‌ها اشاره شده است. Pang و Lee دسته‌بندی احساسات در سطح سند با احساسات مثبت و منفی را تحقیق و بررسی کرده‌اند. آن‌ها با سه الگوریتم مختلف یادگیری ماشین نظیر NB،SVM و ME، آزمایش کرده و فرآیند دسته‌بندی را با استفاده از تکنیک‌های n-gram مانند تک‌واژه‌ای، دوواژه‌ای و ترکیب تک‌واژه‌ای و دوواژه‌ای انجام دادند. آن‌ها چارچوب ویژگی‌های کیفی را برای اجرای الگوریتم‌های یادگیری ماشین استفاده کرده‌اند. درنتیجه تجزیه‌وتحلیل آن‌ها، الگوریتم NB نتایج ضعیف و الگوریتم SVM نتایج بهتری را نشان می‌دهد [۷]. Salvetti و همکارانش، در مفهوم کلی قطبیت عقیده‌ای (OvOp) با استفاده از الگوریتم‌های یادگیری ماشین نظیر مدل بیز ساده و مدل مارکوف را برای دسته‌بندی متون مورد بحث قرار داده‌اند و آن‌ها توسط WordNet و POS متون را برچسب‌گذاری کرده‌اند و

به‌عنوان فیلتر لغوی برای دسته‌بندی عمل می‌کنند. آزمایش آن‌ها نشان می‌دهد که نتیجه به‌دست‌آمده از فیلتر WordNet نسبت به فیلتر POS کمتر است. رویکرد آن‌ها، نتیجه بهتری روی داده‌های وب نشان می‌دهد [۸]. Beineke و همکارانش از مدل بیز ساده برای دسته‌بندی احساسات استفاده کرده‌اند. آن‌ها ویژگی‌های مشتق‌شده را برای پیش‌بینی احساسات استخراج کرده‌اند. آن‌ها ویژگی‌های مشتق شده دیگری را به‌منظور بهبود نتیجه دقت به مدل اضافه کرده‌اند و داده‌های موردنظر را برای تعیین تأثیر نسبی استفاده می‌کنند. این ایده به سیستم اجازه می‌دهد که به‌عنوان یک مدل احتمالی در مقیاس منطقی خطی عمل کند. در این روش هر واژه الحاقی با ضریب k همراه است که این ضرایب باید شناخته شده باشند. بااین‌حال، زمانی که اسناد برچسب‌دار در دسترس هستند، ممکن است تخمین آن‌ها مفید باشد [۹]. Mullen و Collier با استفاده از الگوریتم ماشین بردار پشتیبان ارزش‌ها را به کلمات انتخاب‌شده اختصاص می‌دهند تا یک مدل برای دسته‌بندی ایجاد کنند. علاوه بر این، کلاس‌های مختلفی از ویژگی‌های نزدیک به موضوع اختصاص داده‌شده با مقادیر بهتر به کار برده‌اند که به دسته‌بندی کمک می‌کند. نویسندگان مقایسه‌ای از رویکرد پیشنهادی خود را با داده‌ها، حاشیه‌نویسی موضوع و حاشیه‌نویسی دستی ارائه کرده‌اند. رویکرد پیشنهادی در مقایسه با حاشیه‌نویسی موضوع نشان داده است که در مقایسه با داده‌های حاشیه‌نویسی دستی نتایج بهبود بیشتری دارد. یک مشکل این روش با محدود کردن دامنه با اضافه کردن محدودیت‌های کلمه مرتبط به موضوع این است که تعداد شمارش نتیجه به‌شدت کاهش می‌یابد و هرگونه افزایش بالقوه را از بین می‌برد. هم‌چنین به نظر می‌رسد که جنبه‌های موضوع روابط تحقیق حاضر فقط سطحی بیان می‌شود [۱۰]. Dave و همکارانش تکنیک‌های بازیابی اطلاعات مورداستفاده برای بازیابی ویژگی و نتیجه روش‌های مختلف را مورد آزمایش قرار داده‌اند. آن‌ها از یک ابزار برای سنتز نظرات استفاده کرده‌اند، سپس آن‌ها را تغییر داده و با استفاده از سایت‌های تجمعی دسته‌بندی می‌کنند. آن‌ها با استفاده از این نظرات، ویژگی‌ها را شناسایی کرده و درنهایت روش‌هایی برای تعیین مثبت یا منفی بودن نظرات به کار می‌برند و برای دسته‌بندی جملات حاصل از جستجوی وب با استفاده از نام محصول به‌عنوان شرایط جستجو استفاده می‌کنند. در این روش مواردی چون عدم تطابق درجه‌بندی، هم‌بستگی، مقایسه را دشوار می‌سازد [۱۱]. Matsumoto و همکارانش از ارتباطات نحوی میان کلمات به‌عنوان مبنای تحلیل احساسات در سطح جمله استفاده کرده‌اند. در پژوهش خود، توالی کلمه مکرر و وابستگی جملات را با استفاده از الگوریتم SVM استخراج می‌کنند. آن‌ها روش‌های تک‌واژه‌ای، دوواژه‌ای، توالی کلمه و وابستگی زیردرخت از هر جمله در مجموعه‌داده‌ها را استخراج می‌کنند. این روش در سطح سند امکان‌پذیر نیست [۱۲]. Zhang و Xu و Su روشی برای به دست آوردن ویژگی‌های معنایی با استفاده از word2vec برای گرفتن ویژگی‌های مشابه و سپس دسته‌بندی نظرات از طریق الگوریتم SVM ارائه داده‌اند. رویکرد آن‌ها بر اساس دو بخش است. در بخش اول، آن‌ها از ابزار word2vec برای خوشه‌بندی ویژگی‌های مشابه استفاده کرده‌اند تا ویژگی‌های معناداری در دامنه انتخاب‌شده را ذخیره کنند. سپس در بخش دوم، با انتخاب ویژگی مبتنی بر واژگان به ساخت داده‌های آموزشی می‌پردازند. این روش روی زبان چینی انجام شده است [۱۳]. Liu و Chen چندین دسته‌بند مختلف را در تحلیل احساسات پیشنهاد کرده‌اند. آن‌ها از یازده روش دسته‌بندی چندسطحی در دو مجموعه‌داده‌های میکروبلاگ و هشت ماتریس ارزیابی مختلف برای تجزیه‌وتحلیل استفاده کرده‌اند. آن‌ها هم‌چنین از سه فرهنگ لغت مختلف احساسات برای دسته‌بندی چندسطحی استفاده می‌کنند. این روش از فرهنگ‌نامه از پیش تعریف‌شده استفاده کرده است درحالی‌که می‌توان در دامنه‌های مختلف فرهنگ‌نامه موردنظر را ایجاد کرد و به دلیل برچسب‌های چندگانه مقایسه آزمایش‌های مختلف گران است [۱۴]. Luo و همکارانش روشی برای تبدیل متن به فضای عاطفی کمتر (ESM)، با استفاده از تکنیک‌های یادگیری ماشین جهت دسته‌بندی متون را پیشنهاد کرده‌اند. آن‌ها متن را به کلماتی که معنی قطعی و روشن دارند تبدیل می‌کنند و برای دسته‌بندی کلمات آن را به شش دسته اصلی مانند خشم، ترس، انزجار، غم و اندوه، شادی و تعجب تقسیم می‌کنند. آن‌ها دو روش متفاوت برای تخصیص وزن به کلمات را با برچسب‌های عاطفی در نظر گرفته‌اند و تمام وزن کلمات عاطفی را محاسبه کرده‌اند و بر اساس این مقادیر، پیام‌ها به گروه‌های مختلف دسته‌بندی می‌شوند. این روش می‌تواند نتایج منطقی را در هر مجموعه‌داده یا دامنه استفاده کند ولی تبدیل کلمات به معنای قطعی و تقسیم آن‌ها به شش دسته اصلی زمان‌بر است [۱۵]. Niu

و همکارانش مجموعه‌ای از داده‌ها را که حاوی مجموعه‌ای از جفت‌متن تصویر با حاشیه‌نویسی دستی از توییتر است جمع‌آوری کرده‌اند. رویکرد آن‌ها از تجزیه‌وتحلیل احساسات دارای دو بخش یادگیری مبتنی بر واژگان و آماری است که در مورد تجزیه‌وتحلیل مبتنی بر واژگان، مجموعه‌ای از کلمات یا عبارات در نظر گرفته شده است و نمره احساسات از پیش تعریف‌شده دارند. درحالی‌که در یادگیری آماری، تکنیک‌های مختلف یادگیری ماشین با ویژگی‌های متن مورد استفاده قرار می‌گیرد [۱۶]. Caiqiang و Junming مدل ارزیابی معلم آموزشی آنلاین را بر اساس نظرکاوی انجام می‌دهند. آن‌ها با استفاده از خزنده وب، نظرات دانش‌آموزان را که در زبان چینی منتشر شده و در سیستم مدیریت یادگیری نوشته شده است، جمع‌آوری می‌کنند. اگر کلمات احساسی متن ذهنی در لغت‌نامه قطب‌واژه نیستند، آن‌ها از روش متقابل اطلاعات متناوب برای قضاوت قطبیت استفاده کردند. این مدل یک ارزیابی کلی از هر معلم ارائه می‌دهد [۱۷]. Ortigosa، Martin و Carro روش دیگر تحلیل احساسات برای محیط یادگیری الکترونیکی را با استفاده از روش ترکیبی از تکنیک‌های یادگیری زبان اسپانیایی پیشنهاد دادند. در آزمایش آن‌ها ترکیبی از تکنیک‌های مبتنی بر واژگان و SVM بالاترین دقت را به دست می‌آورند. در زمینه یادگیری الکترونیکی، ممکن است اطلاعاتی در مورد احساسات دانش‌آموز از پیام‌هایی که در فیس‌بوک می‌نویسند، استخراج شود. احساسات دانش‌آموزان نسبت به یک دوره می‌تواند به‌عنوان بازخورد برای معلمان، به‌ویژه در مورد یادگیری آنلاین، مفید باشد. بااین‌حال، این کار همچنان دارای محدودیتی برای تحلیل احساسات است، تمام کلمات به‌عنوان یک قطب، نظرات را برچسب‌گذاری می‌کنند، تمام کلمات مثبت نمره ۱، تمام کلمات منفی نمره ۱- و تمام کلمات خنثی نمره صفر را می‌گیرند آن‌ها وزن‌های مختلف را به کلمات مختلف اختصاص نمی‌دهند و برای بدتر و بد امتیاز مشابه به دست می‌آورند [۱۸]. Pong-inwong و Rungworawut پیشنهاد ساخت تدریس تئوری ارزیابی احساسات را برای قطبیت احساسات خودکار ارائه دادند. کار آن‌ها متشکل از دو بخش است، بخش اول تهیه داده‌ها و بخش دوم مدل‌سازی داده‌ها و ارزیابی است. نویسندگان انتخاب ID3، بیزین و ماشین بردار پشتیبان را برای دسته‌بندی ارزیابی احساسات آموزشی انتخاب کردند. نتایج تجربی آن‌ها نشان می‌دهد که SVM دارای بالاترین دقت است. در این کار، نمره وزن توسط یک متخصص تعریف شده است که دارای تجربه در ارزیابی تدریس است. وزن احساسات از ۱.۰۰- به ۱.۰۰+ متغیر بود. روش پیشنهادی آن‌ها نمی‌تواند واژه تقویت‌کننده را در نظر بگیرد و در زبان تایلندی ساخته شده است [۱۹]. Yili Wang و Hee Yong Youn روش وزن‌دهی تطبیقی جدید به نام میان‌رده قدرت تشخیصی با مدل Chi-square پیشنهاد کرده‌اند که مقیاس صحیح میان‌رده و اهمیت داخل دسته‌ای را مشخص می‌کند. یک مدل ریاضی جدید برای اندازه‌گیری قدرت تشخیصی میان‌رده‌ها از ویژگی‌ها ارائه کرده‌اند، درحالی‌که یک مدل Chi-square اصلاح‌شده برای اندازه‌گیری وابستگی ویژگی‌ها در داخل دسته‌بندی ارائه شده است. هم‌چنین استراتژی خوشه‌بندی ریزنمونه برای تعیین حاکمیت ویژگی‌های تشخیصی و تعیین ویژگی‌های توزیع مشابه برای وزن‌دهی کارایی پیشنهاد شده است. علاوه بر این، یک استراتژی وزن سازگاری پیشنهاد شده است تا وزن هر ویژگی را به‌درستی تعیین کند. به‌طور خاص، توزیع یک کلمه در یک دسته خاص نشان‌دهنده وابستگی آن به آن دسته است که با توجه به وابستگی وزن، دقیق‌تر می‌تواند باشد. ازآنجاکه طرح پیشنهادی بر اساس وابستگی هر یک از ویژگی به رده مربوطه است، طول مدت جمله به‌طور مستقیم بر عملکرد روش پیشنهادی تأثیر نمی‌گذارد و همان‌طور که هر کلمه دارای وزن مشخصی است منعکس‌کننده درجه اهمیت آن برای دسته‌های مختلف است، قطبیت یک جمله دارای دو دیدگاه مختلف است که توسط وزن‌های به‌دست‌آمده از دسته‌بندی متن تعیین می‌شود. برای عقاید مختلف یا متضاد ایجاد نشده است. در طرح پیشنهادی شکلک‌ها با استفاده از تابعی در نرم‌افزار متلب شناسایی و حذف می‌شوند درحالی‌که احساسات به‌طورمعمول احساسات مردم را در برمی‌گیرند و این روش مناسب نیست [۲۰].

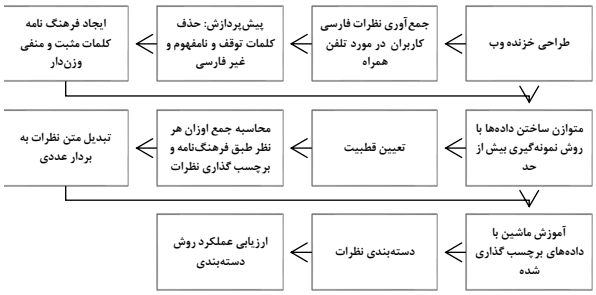
۳- رویکرد پیشنهادی

داده‌کاوی یکی از گرایش‌های بسیار مهم در علم کامپیوتر است. در داده‌کاوی، سعی بر این است که دانش از حجم انبوهی از داده‌ها استخراج شود. منظور از دانش، اطلاعات مطلوبی است که مفید و هدفمند و از قبل شناخته شده- اند. در این میان نظرکاوی رشته‌ای است که به‌تازگی بسیار موردتوجه قرار گرفته است و تحقیقات زیادی برای بررسی و توسعه تکنیک‌های مربوط به نظرکاوی انجام شده است. نظرکاوی قانونی در تقاطع بازیابی اطلاعات و محاسبات

زبان‌شناسی است که نظرات کاربران در مورد یک موضوع داده‌شده را کاوش می‌کند. هدف نظرکاوی این است که نظر نویسنده و یا یک نظر کلی در مورد یک سند خاص را استخراج کند و راهکارهایی برای دسته‌بندی نظرات مندرج در سایت‌ها و وبلاگ‌ها است. در حقیقت در نظرکاوی به دنبال یافتن تابعی هستیم که بتواند با دریافت ورودی یک کامنت آن را از منظر احساسات به یکی از مقادیر مثبت و منفی نگاشت نماید. این مسئله در معادله ۱ فرموله شده است.

$$Label = f_{opinion-mining}(Comment)$$
$$Label \in \{Pos, Neg\} \quad (1)$$

که در آن $Comment$ نظر کاربر و $Label$ نتیجه نظرکاوی شامل یکی از مقادیر مثبت و منفی خواهد بود. ازآنجاکه برآورد همچنین تابعی ازنظر روش‌های مرسوم ریاضی امکان‌پذیر نخواهد بود، لذا استفاده از الگوریتم‌های یادگیری ماشین جهت تقریب تابع فوق مفید خواهد بود. چنانچه بخواهیم از الگوریتم‌های یادگیری ماشین برای این منظور استفاده کنیم نیاز به یک مجموعه‌داده آموزشی خواهیم داشت که برچسب گذاری شده باشد. مراحل رویکرد پیشنهادی در (شکل(۱) مورد بحث قرار گرفته است.



(شکل – ۱). نمایی از روش پیشنهادی
(Figure - 1). Proposed Method Schema

ابتدا با استفاده از یک خزنده وب طراحی‌شده در زبان پایتون نظرات فارسی کاربران در مورد انواع تلفن همراه استخراج و در یک مجموعه نظرات غیر ساخت‌یافته ذخیره می‌گردد. مجموعه نظرات غیر ساخت‌یافته استخراج‌شده در معادله ۲ فرموله شده است.

$$UnCmts = \{Cmt_1, Cmt_2, \dots, Cmt_n\} \quad (2)$$

که در آن Cmt_i نظر کاربر i–ام و n تعداد کل نظرات استخراج شده است.

در مرحله پیش‌پردازش، برای بهبود تعیین قطبیت (مثبت و منفی) یک نظر، داده‌های اضافی از نظرات، مانند کلمات پایانی و کلمات غیرکلیدی مانند از، اگر، به و غیره، اعداد، علائم نقطه‌گذاری، حروف انگلیسی، حروف و علائم غیرفارسی، حذف فاصله‌ها، حروف تکراری بیش از دو بار در یک کلمه، کلماتی با طول یک که معنی خاصی ندارند و نمی‌توان برای آن‌ها قطبیت تعیین کرد را از متن اصلی برای بهبود نظرکاوی و تعیین دقیق قطبیت حذف کرده‌ایم که این عملیات را پیش‌پردازش متن می‌نامیم.

به‌طور خلاصه با توجه به الگوریتم (۱) ارائه‌شده برای پیش‌پردازش متون نظرات تلفن همراه عملیات زیر انجام شده است:

- حذف اعداد و علائم و کلمات غیر حروف فارسی مانند علائم نقطه‌گذاری، تعجب، اصطلاحات انگلیسی و غیره (خط ۲، ۳ و ۴).
- حذف حروف تکراری بیش از یک‌بار در کلمه به‌طور مثال به‌جای کلماتی چون عالللللی کلمه عالی جایگزین می‌شود (خط ۵).
- حذف کلمات یک‌حرفی که در جمله بی‌معنا هستند (خط ۶).
- نرمال‌سازی متن که به‌منظور حذف فاصله‌ها است (خط ۷).
- جداسازی کلمات به‌منظور استخراج ویژگی‌ها استفاده می‌شود (خط ۸).
- حذف کلمات پایانی مانند اگر، به، چون و غیره که آن‌ها در تعیین احساسات نقش مهمی ایفا نمی‌کنند (خط ۹).

ALGORITHM Preprocessing Pseudo Code

INPUT: Unstructured Comments (*UnCmts*)
Dataset of the User's Comments
OUTPUT: Structural Comments ($\mathcal{C}$)
Structured Dataset of the User's Comments

1. FOR EACH *c* IN *UnCmts*
2. Remove Digits
3. Remove Punctuation Mark
4. Remove Non-Persian letter
5. Remove Repeat Characters
6. Remove Repeat Words
7. Remove Spaces (Normalization)
8. Tokenize Words
9. Remove Persian StopWords
10. Add *c* To Structured Comments $\mathcal{C}$
11. Return $\mathcal{C}$

(الگوریتم-۱). پیش پردازش

مجموعه اعمال مربوط به پیش‌پردازش نظرات در معادله ۳ فرموله شده است.

$$\mathcal{C} = Preprocess(UnCmts)$$
$$\mathcal{C} = \{C_1, C_2, \dots, C_n\} \quad (3)$$
$$C_i = \{kw_{1_i}, kw_{2_i}, \dots, kw_{m_i}\}$$

که در آن C مجموعه نظرات ساختاریافته و n تعداد نظرات و kw_i کلمات کلیدی حذف نشده ازنظر i–ام و m_i تعداد کلمات کلیدی نظر i–ام هستند. در تجزیه‌وتحلیل

احساسات، مدل چندواژه‌ای مانند تک‌واژه‌ای، دوواژه‌ای، سه‌واژه‌ای و برای نمونه‌های بزرگ‌تر چهارواژه‌ای و پنج-واژه‌ای به تجزیه‌وتحلیل احساسات متن یا سند کمک می-کند. به‌عنوان‌مثال روش‌های چندواژه‌ای را با استفاده ازجمله "The movie is not a good one" شرح می‌دهیم. در روش تک‌واژه‌ای جمله فوق به‌صورت تک کلمه به شرح 'The' و 'movie' و 'is' و 'not' و 'a' و 'good' و 'one' است. در روش دوواژه‌ای جفت کلمات در نظر گرفته شده است. بنابراین کلمات کلیدی جمله فوق در این روش به شرح 'The movie' و 'movie is' و 'is not' و 'not a' و 'a good' و 'good one' است. در روش سه‌واژه‌ای مجموعه‌ای از کلمات سه‌تایی در نظر گرفته می‌شود. بنابراین کلمات کلیدی جمله فوق در این روش به شرح 'The movie is' و 'movie is not' و 'is not a' و 'not a good' و 'a good one' است.

بیشتر نویسندگان از روش تک‌واژه‌ای برای طبقه‌بندی نظرات استفاده کرده‌اند. این رویکرد نتیجه‌ای بهتر را فراهم می‌کند چون وابستگی به کلمات قبل خود ندارد، اما در بعضی موارد شکست می‌خورد به‌عنوان‌مثال در جمله"The movie is not a good one" هنگامی‌که با استفاده از روش تک‌واژه‌ای تجزیه‌وتحلیل می‌شود، قطبیت جمله باوجود یک قطب مثبت کلمه "good" و یک قطب منفی کلمه "not" خنثی می‌شود؛ اما هنگامی‌که این جمله با استفاده از رویکرد دوواژه‌ای تجزیه‌وتحلیل می‌شود، قطبیت جمله به دلیل وجود کلمات "not good " منفی تشخیص داده می‌شود؛ بنابراین، انتظار می‌رود که نتیجه بهتر باشد.

پس از پیش‌پردازش نظرات، از کلمات کلیدی باقی‌مانده در کل داده‌ها، یک فرهنگ‌نامه به‌صورت دستی ایجاد می‌کنیم و به هر یک از کلمات با توجه به کاربرد آن‌ها امتیازی بین ۵- تا ۵+ می‌دهیم. درواقع کلمات خیلی ضعیف منفی با نمره (۵-)، کلمات ضعیف منفی با نمره (۴-)، خصوصیات منفی تلفن همراه با نمره (۳-)، کلمات متوسط منفی با نمره (۲-)، کلمات تقریباً نادرست منفی با نمره (۱-)، کلمات نامفهوم با نمره (۰)، کلمات تقریباً نادرست مثبت با نمره (۱)، کلمات متوسط مثبت با نمره (۲)، خصوصیات مثبت تلفن همراه با نمره (۳)، کلمات قوی مثبت با نمره (۴) و کلمات خیلی قوی مثبت با نمره (۵) امتیازدهی می‌شوند. ساخت فرهنگ‌نامه در معادله ۴ فرموله شده است.

$$Dic = \bigcup_{i=1}^{n} C_i = \{all\ unique\ kw\ in\ \mathcal{C}\}$$
$$Score: Dic \rightarrow \{-5, -4, \dots, +4, +5\}$$
$$for\ example \qquad (۴)$$
$$Score\left(\text{عالی}\right) = +5, Score\left(\text{افتضاح}\right) = -5$$

که در آن Dic فرهنگ‌نامه‌ای شامل اجتماع کلیه لغات کلیدی در نظرات کاربران است و تابع $Score$ به ازای هر لغت کلیدی عددی بین ۵+ و ۵- را نگاشت می‌کند. نبود یک فرهنگ‌نامه وزن‌دهی به کلمات در زبان فارسی ازلحاظ بار احساسی به‌ویژه در زمینه تلفن همراه از چالش‌های مهم این پژوهش به شمار می‌آید. این فرهنگ‌نامه بیشتر لغات استفاده‌شده توسط کاربران در نظرسنجی تلفن همراه را جمع‌آوری و ازلحاظ بار احساسی امتیازدهی نموده است.

مرحله بعدی تعیین قطبیت هر یک از نظرات کاربران و برچسب گذاری آن‌ها جهت ایجاد داده آموزشی است. این کار با استفاده از فرهنگ‌نامه ساخته‌شده در مرحله قبل صورت می‌پذیرد. الگوریتم (۲) جزییات برچسب‌گذاری را بیان می‌کند.

ALGORITHM Labeling Pseudo Code

INPUT: Structural Comments ($\mathcal{C}$)
Structured Dataset of the User's Comments
Keyword's Dictionary (Dic)
Score function
OUTPUT: Labeled User's Comments ($\mathcal{L}$)

1. FOR EACH $C\ in\ \mathcal{C}$
2. FOR EACH $kw\ in\ C$
3. S = Sum all $Score(kw)$
4. IF $S > 0$ THEN $L = POS$
5. IF $S < 0$ THEN $L = NEG$
6. ELSE $remove\ C\ from\ \mathcal{C}$
7. Add $(C, L)\ to\ \mathcal{L}$
8. Return $\mathcal{L}$

(الگوریتم-۲).برچسب گذاری

در این مرحله طبق الگوریتم برچسب‌گذاری، هر نظر را به‌طور جداگانه (خط ۱) به‌صورت کلمه به کلمه تفکیک کرده (خط ۲) و امتیاز هر کلمه را از فرهنگ‌نامه ساخته-شده در مرحله قبل (خط ۳) پیدا می‌کنیم. اگر کلمات موجود در نظر در فرهنگ‌نامه وجود داشت، امتیاز آن کلمات را با هم جمع می‌کنیم (خط ۳). اگر مجموع امتیازات یک نظر بیشتر از صفر باشد آن نظر برچسب مثبت (خط ۴) و اگر کمتر از صفر باشد برچسب منفی را به خود اختصاص می‌دهد (خط ۵)؛ بنابراین کل مجموعه‌داده به این روش برچسب‌گذاری می‌شوند.

روال برچسب‌گذاری در معادله ۵ فرموله شده است. $\mathcal{L}$ مجموعه زوج مرتب‌های (نظر، برچسب) است به‌طوری‌که چنانچه مجموع امتیازات کلمات کلیدی هر نظر بزرگ‌تر از صفر شود برچسب مثبت و اگر کوچک‌تر از صفر شود برچسب منفی خواهد خورد. همچنین m_C تعداد کلمات کلیدی هر نظر است.

$$\mathcal{L} = \{(C, L) \mid C \in \mathcal{C}, L \in \{POS, NEG\}\} \qquad (۵)$$

$$L = \begin{cases} POS & \sum_{i=1}^{m_c} Score(kw_i) > 0 \\ NEG & \sum_{i=1}^{m_c} Score(kw_i) < 0 \end{cases}$$

چالش مهم بعدی در نظرات استخراج‌شده از سایت‌ها، عدم توازن بین نظرات مثبت و منفی است که می‌تواند بر نتیجه الگوریتم یادگیری تأثیر منفی داشته باشد. به‌دلایل مختلف مثل بازاریابی، تبلیغات، مسائل فرهنگی، وجود الفاظ نامناسب و ... بسیاری از نظرات منفی از سایت‌ها حذف می‌شوند و درنتیجه تعداد نظرات منفی استخراج‌شده عموماً کمتر از نظرات مثبت است. ازاین‌رو از روش‌های متعادل‌سازی، ازجمله روش نمونه‌گیری بیش‌ازحد و روش نمونه‌گیری کمتر از حد برای بهبود عملکرد دسته‌بندی می‌توان استفاده نمود. در این پژوهش از روش نمونه‌گیری بیش‌ازحد استفاده شده است زیرا این روش برای مجموعه‌داده‌های کوچک مناسب‌تر است و مناسب‌ترین استراتژی جهت افزایش مجموعه نمونه اولیه است و این کار را با تمرکز بر روی نمونه‌های دودویی یا چند ستونی انجام می‌دهد. در این روش نمونه‌های بیشتری به «کلاس کم‌جمعیت‌تر» اضافه می‌شوند. البته در تولید نمونه‌های جدید سعی می‌شود بیشترین شباهت با نمونه‌های موجود صورت پذیرد تا توزیع آماری نمونه‌ها تغییر نکند. با توجه به الگوریتم ۳ متوازن‌سازی نمونه‌های جدید با ترکیب تصادفی نمونه‌های موجود در داده‌ها با برچسب منفی ساخته می‌شوند (خط ۳-۵). این کار آن‌قدر ادامه پیدا می‌کند تا تعداد نمونه‌های هر دو کلاس نسبتاً برابر شوند (خط ۶).

ALGORITHM Balancing Pseudo Code

INPUT: Labeled User's Comments ($\mathcal{L}$)
OUTPUT: Balanced Labeled User's Comments ($\mathcal{L}^*$)

1. $\mathcal{L}^* = \mathcal{L}$
2. REPEAT
3. Choose two random C_1, C_2 $from\ \mathcal{L}\ where\ L = NEG$
4. $C^* = Combination\ of\ C_1, C_2$
5. Add (C^*, NEG) $to\ \mathcal{L}^*$
6. UNTIL $|[POS]| = |[NEG]|$
7. Return $\mathcal{L}^*$

(الگوریتم-۳). متوازن‌سازی

روال متوازن‌سازی در معادله ۶ فرموله شده است.

$$\mathcal{L}^* = \mathcal{L} \cup \{(C^*, NEG)| \\ C^*: Mixture\ of\ 2\ random\ neg\ comments\} \quad (۶)$$

در این مرحله برای استفاده از الگوریتم‌های یادگیری ماشین و یادگیری عمیق برای ارزیابی عملکرد روش پیشنهادی داده‌های متنی باید به بردارهای عددی تبدیل شوند. در این تحقیق با استفاده از روش بردار ویژگی که اهمیت کلمه را به سند نشان می‌دهد برای ایجاد بردار عددی استفاده شده است. بردارهای عددی حاصل از مجموعه نظرات تشکیل ماتریس $\mathcal{T}$ را می‌دهند. طبق الگوریتم (۴) بردارسازی ابتدا برای هر نظر مقدار فراوانی کلمات کلیدی آن محاسبه می‌شود. سپس این مقدار با تقسیم بر تعداد کل فراوانی نرمال می‌گردد (خط ۱). مقدار اهمیت هر کلمه در کل نظرات نیز با محاسبه لگاریتم معکوس فراوانی آن کلمه در کل نظرات تقسیم بر فراوانی در آن نظر به دست می‌آید (خط ۲). ضرب این دو مقدار میزان اهمیت یک کلمه در یک نظر را مشخص می‌کند. به‌این‌ترتیب کلماتی که در بیشتر نظرات وجود دارند اهمیت تصمیم‌گیری کمتری پیدا می‌کنند. در انتها به ازای همه نظرات و همه کلمات موجود در فرهنگ‌نامه مقدار اهمیت در ماتریس $\mathcal{T}$ ذخیره می‌گردد (خط۳-۶). این ماتریس عددی به‌عنوان داده آموزشی و آزمون برای دسته‌بندها مورد استفاده قرار خواهد گرفت.

ALGORITHM Vectorizer Pseudo Code

INPUT: Balanced Labeled User's Comments ($\mathcal{L}^*$)
Structural Comments ($\mathcal{C}$)
Keyword's Dictionary (Dic)
OUTPUT: Labeled TF/IDF Matrix ($\mathcal{T}$)

1. TF(kw ,C): Return $\left(\frac{freq(kw)}{number\ of\ words\ in\ C}\right)$
2. IDF(kw, $\mathcal{C}$):
Return $log\left(\frac{number\ of\ comments\ |\mathcal{C}|}{number\ of\ comments\ including\ kw}\right)$
3. FOR EACH (C, L) $in\ \mathcal{L}^*$
4. FOR EACH kw $in\ Dic$
5. $\mathcal{T}[C, kw] = TF(kw, C) \times IDF(kw, \mathcal{C})$
6. $\mathcal{T}[C, class] = L$
7. Return $\mathcal{T}$

(الگوریتم-۴). بردارسازی

مراحل ساخت ماتریس در معادله ۷ فرموله شده است؛ که در آن $freq$ نمایانگر فراوانی یک کلمه در سند است. بقیه متغیرها مانند معادلات قبل هستند.

$$TF(kw, C) = \left(\frac{freq(kw)}{\sum_{kw^* \in C} freq(kw^*)}\right)$$
$$IDF(kw, \mathcal{C}) = log\left(\frac{|\mathcal{C}|}{|\{C \in \mathcal{C} : kw \in C\}|}\right)$$
$$\mathcal{T}_{n \times (m+1)}: Matrix\ of\ (n = |\mathcal{L}^*|) \times (m + 1 = |Dic| + 1)\ elements \quad (۷)$$
$$\forall (L_i, C_i) \in \mathcal{L}^*\ \ i = 1..n$$
$$\forall kw_j \in Dic\ \ j = 1..m:$$
$$\mathcal{T}[i, j] = TF(kw_j, C_i) \times IDF(kw_j, \mathcal{C})$$
$$\mathcal{T}[i, m+1] = L_i$$

پس از ساخت ماتریس داده‌ها نوبت به دسته‌بندی داده‌ها با الگوریتم‌های یادگیری ماشین است الگوریتم (۵). به این منظور داده‌ها به دو بخش آموزشی (۷۰٪) و آزمون (۳۰٪) به‌صورت تصادفی تقسیم می‌شوند (خط ۲). این ماتریس‌ها طبق الگوریتم دسته‌بندی با پنج الگوریتم مختلف یادگیری ماشین تحت نظارت ازجمله بیز ساده (خط ۳)، ماشین بردار پشتیبان (خط ۴)، کاهش گرادیان تصادفی (خط ۵)، رگرسیون لجستیک (خط ۶) و جنگل

تصادفی (خط ۷) و هم‌چنین از الگوریتم‌های یادگیری عمیق مانند شبکه عصبی کانولوشن (خط ۸) و حافظه کوتاه‌مدت متوالی (خط ۹) یادگیری و مقادیر دقت، صحت، بازیابی و معیار فیشر هرکدام ذخیره می‌گردد. برای دستیابی به دقت بالاتر و از بین بردن اثر بیش‌برازش این کار به تعداد دفعات مختلف اجرا و میانگین نتایج جهت ارزیابی مورد استفاده قرار می‌گیرد (خط ۱۰).

ALGORITHM Classification Pseudo Code

INPUT: Labeled TF/IDF Matrix ($\mathcal{T}$)
OUTPUT: Accuracy,Precision,Recall,F-Measure
DeepAccuracy, DeepLoss

1. Repeat NumFolds
2. [Train,Test]=Random Split of $\mathcal{T}$ (70%,30%)
3. [Accuracy,Precision,Recall,F-Measure] = NB(Train,Test)
4. [Accuracy,Precision,Recall,F-Measure] = SVM(Train,Test)
5. [Accuracy,Precision,Recall,F-Measure] = SGD(Train,Test)
6. [Accuracy,Precision,Recall,F-Measure] = LR(Train,Test)
7. [Accuracy,Precision,Recall,F-Measure] = RF(Train,Test)
8. [DeepAccuracy, DeepLoss] = CNN($\mathcal{T}$)
9. [DeepAccuracy, DeepLoss] = LSTM($\mathcal{T}$)
10. [Accuracy,Precision,Recall,F-Measure]/NumFolds
11. Return [Accuracy,Precision,Recall,F-Measure]
12. Return [DeepAccuracy,DeepLoss]

(الگوریتم-۵). دسته‌بندی

مفروضات زیر جهت کلیه دسته‌بندها برقرار است.

$$Train\ the\ Classifier\ using\ Training\ Data\ \mathcal{T}$$
$$\mathcal{T} = \{\mathcal{T}_i\}_{i=1..n}, \mathcal{T}_i = (\boldsymbol{T}_i, L_i), \boldsymbol{T}_i = \left[TF.IDF(kw_j)\right]_{j=1..m}$$
$$L_i \in \{pos, neg\}$$
$$Classifier\ F: \boldsymbol{T} \rightarrow L$$

که در آن بردار $\boldsymbol{T}$ به‌عنوان ورودی دسته‌بند شامل مقادیر TF.IDF کلمات کلیدی یک نظر و برچسب L به‌عنوان خروجی تابع دسته‌بندی در نظر گرفته شده است. به ازای نظرات جدید (دیده نشده) دسته‌بند آموزش‌دیده سعی در پیش‌بینی برچسب می‌کند.

$$for\ a\ given\ comment\ (Cmt):$$
$$C = Preprocessing(Cmt)\ \ s.t.\ \ C = \{kw_j\}_{j=1..m_C}$$
$$\boldsymbol{T} = Vectorizer(C) = \left[TF.IDF(kw_j)\right]_{j=1..m}$$
$$predict\ \hat{L} = F_\theta(\boldsymbol{T})$$
$$Error = |L - \hat{L}|$$

بدین ترتیب که ابتدا نظر پس از پیش‌پردازش به مجموعه‌ای از کلمات کلیدی تبدیل و سپس مقادیر TF.IDF کلمات محاسبه و در بردار $\boldsymbol{T}$ ذخیره می‌شود. مقدار قطبیت نظر توسط دسته‌بند آموزش‌دیده‌شده محاسبه می‌گردد $(\hat{L})$. این مقدار بعداً جهت ارزیابی دقت و صحت الگوریتم قابل‌استفاده است.

در بخش بعدی به ارزیابی رویکرد پیشنهادی روی دو مجموعه‌داده نظرات جمع‌آوری‌شده از سایت دیجی‌کالا در مورد تلفن همراه و داده‌های سنتی‌پرس می‌پردازیم.

۴- ارزیابی عملکرد رویکرد پیشنهادی

در این بخش ابتدا مجموعه‌داده‌های استفاده‌شده، سپس معیارهای ارزیابی و درنهایت سناریوهای مختلف ارزیابی رویکرد پیشنهادی مورد بررسی قرار گرفته است. کلیه ارزیابی‌ها توسط زبان برنامه‌نویسی پایتون انجام شده است [۲۶].

۴-۱- مجموعه‌داده

در این پژوهش، مجموعه‌داده‌های فارسی نظرات تلفن همراه از سایت دیجی‌کالا و مجموعه‌داده‌های سنتی‌پرس از سایت پیکره‌گان برای تحلیل احساسات در نظر گرفته شده است. این مجموعه‌داده دیجی‌کالا را با استفاده از خزنده وب جمع‌آوری کرده‌ایم که شامل ۳۴۶۲۵ نظرسنجی با برچسب مثبت و ۵۰۸۵ نظرسنجی با برچسب منفی است. هم‌چنین مجموعه‌داده سنتی‌پرس از سایت پیکره‌گان قابل دریافت است که شامل ۷۴۱۰ نظرسنجی با برچسب مثبت و ۲۴۹۵ نظرسنجی با برچسب منفی و ۳۴۱۵ نظرسنجی با برچسب خنثی است. در این پژوهش برای بهبودی نظرکاوی یک فرهنگ‌نامه از کلمات مثبت و منفی داده‌های نظرات تلفن همراه ایجاد کرده‌ایم و به هر یک از کلمات با توجه به ارزش آن‌ها وزنی اختصاص داده‌ایم. این فرهنگ‌نامه برای داده‌های دیجی‌کالا شامل ۱۴۰۹۰ لغت با وزن‌های مختلف از ۵- تا ۵+ است که دارای ۷۲۷۰ کلمه با خصوصیت مثبت و ۶۸۲۰ کلمه با خصوصیت منفی است. مجموعه‌داده‌های دیجی‌کالا بعد از به دست آوردن قطبیت جدید نظرات طبق فرهنگ‌نامه ارائه‌شده دارای ۳۱۸۹۵ نظرسنجی با برچسب مثبت و ۶۶۱۵ نظرسنجی با برچسب منفی است. هم‌چنین فرهنگ‌نامه برای داده‌های سنتی‌پرس شامل ۱۴۲۰۵ لغت با وزن‌های مختلف از ۵- تا ۵+ است که دارای ۳۳۱۵ کلمه با خصوصیت مثبت و ۱۳۱۰ کلمه با خصوصیت منفی و ۹۵۸۰ کلمه نامفهوم است. مجموعه‌داده‌های سنتی‌پرس بعد از به دست آوردن قطبیت جدید نظرات طبق فرهنگ‌نامه ارائه‌شده دارای ۷۴۱۰ نظرسنجی با برچسب مثبت و ۲۴۹۵ نظرسنجی با برچسب منفی است. سپس به‌دلیل ناهمگن‌بودن داده‌های مثبت و منفی با استفاده از روش نمونه‌گیری بیش‌ازحد، کل داده‌ها را به بالاترین مقدار سوق می‌دهیم و داده‌های

دیجی‌کالا با برچسب مثبت و منفی هر یک دارای ۳۱۸۹۵ و داده‌های سنتی‌پرس با برچسب مثبت و منفی هر یک دارای ۷۴۱۰ نظرسنجی می‌شوند. بعد از آماده‌سازی داده‌ها از ۷۰ درصد آن‌ها برای آموزش و ۳۰ درصد برای آزمایش استفاده کرده‌ایم. داده‌های آزمایش از روش ۱۰ دور استفاده می‌کند یعنی داده‌های آزمایش در ۱۰ جای مختلف از کل داده آزمایش می‌شود و میانگین دقت را برمی‌گرداند.

۴-۲- پارامترهای ارزیابی عملکرد

پارامترهای مفید برای ارزیابی عملکرد الگوریتم یادگیری ماشین تحت نظارت از یک ماتریس شناخته‌شده به‌عنوان ماتریس درهم‌ریختگی یا جدول احتمالی استخراج می‌شوند. این پارامترها در الگوریتم یادگیری ماشین تحت نظارت مورد استفاده قرار می‌گیرند که در ارزیابی عملکرد هر الگوریتم کمک می‌کند. ازنظر دسته‌بندی، برای مقایسه برچسب کلاس‌ها اصطلاحات مثبت واقعی (TP)، مثبت کاذب (FP)، منفی واقعی (TN)، منفی کاذب (FN) محاسبه می‌شوند. مثبت واقعی، تعداد نظراتی را نشان می‌دهد که مثبت هستند و هم‌چنین در دسته‌بندی به‌صورت مثبت دسته‌بندی می‌شوند، مثبت کاذب نشان‌دهنده نظرات مثبت است اما دسته‌بندی آن را به‌صورت مثبت دسته‌بندی نمی‌کند. به‌طور مشابه، منفی واقعی نشان‌دهنده نظرات منفی است که توسط دسته‌بندی منفی دسته‌بندی شده‌اند و منفی کاذب، نظرات منفی هستند اما دسته‌بندی آن به‌صورت منفی دسته‌بندی نمی‌شود [۳۶]. این مقادیر در **Error! Reference source not found.**(۱) نشان داده شده است.

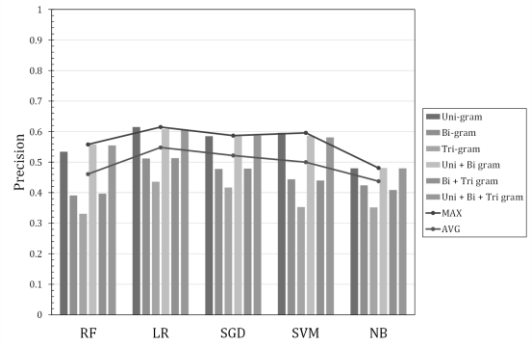
بر اساس مقادیر به‌دست‌آمده از جدول فراوانی، برای ارزیابی عملکرد هر دسته‌بندی، پارامترهای دقت، صحت، بازیابی و معیار فیشر به دست می‌آیند. دقت رایج‌ترین پارامتر فرآیند دسته‌بندی است که می‌توان به‌عنوان نسبت نمونه‌های دسته‌بندی‌شده مثبت به تعداد کل نمونه‌ها محاسبه شود. صحت دقیق بودن نتیجه دسته‌بندی را اندازه‌گیری می‌کند و نسبت تعداد نمونه‌هایی است که به‌طور صحیح، مثبت برچسب‌گذاری شده‌اند به تعداد کل نمونه‌هایی که مثبت دسته‌بندی شده‌اند. بازیابی کامل بودن نتیجه دسته‌بندی را اندازه‌گیری می‌کند و نسبت تعداد کل نمونه‌های مثبت برچسب‌گذاری شده به نمونه‌های کاملی است که واقعاً مثبت هستند. معیار فیشر هارمونیک متوسط صحت و بازیابی است. لازم است که سیستم را به سمت صحت یا بازیابی بهینه‌سازی کنیم که تأثیر بیشتری بر نتیجه نهایی دارد. [۳۶]

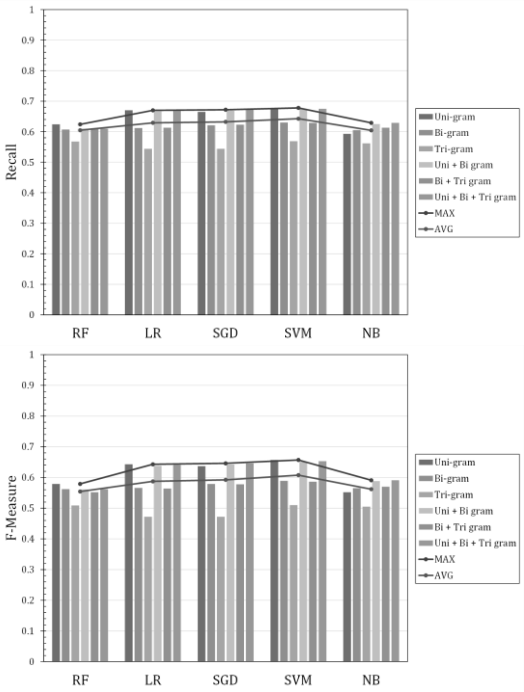
(جدول - ۱). فراوانی نتایج صحیح و اشتباه دسته‌بندی
(Table - 1). True/False Classification result frequency

برچسب‌گذاری / دسته‌بندی	مثبت	منفی
مثبت	مثبت واقعی (TP)	مثبت کاذب (FP)
منفی	منفی واقعی (TN)	منفی کاذب (FN)
دقت	$\text{Accuracy} = \frac{TP + \text{TN}}{TP + \text{TN} + FP + \text{FN}}$	
صحت	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$	
بازیابی	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	
معیار فیشر	$\text{F} - \text{Measure} = \frac{2 * \text{Precision} * \text{Recall}}{Precesion + Recall}$	

۴-۳- تأثیر استفاده از فرهنگ‌نامه و متوازن‌سازی بر داده‌های دیجی‌کالا

در این قسمت نتایج پارامترهای ارزیابی نظرات تلفن همراه از سایت دیجی‌کالا با استفاده از الگوریتم‌های یادگیری ماشین توسط تکنیک‌های تک‌واژه‌ای، دوواژه‌ای، سه‌واژه‌ای، ترکیب تک‌واژه‌ای و دوواژه‌ای، ترکیب دوواژه‌ای و سه‌واژه‌ای و ترکیب هر سه تکنیک بررسی شده است. با استفاده از ماتریس درهم‌ریختگی، پارامترهای مختلف ارزیابی مانند دقت، صحت، بازیابی و معیار فیشر به‌دست‌آمده پس از اعمال روش پیشنهادی محاسبه شده است. دسته‌بندی نتایج بر اساس انواع تکنیک‌های چندواژه‌ای و با استفاده از الگوریتم‌های مختلف یادگیری ماشین در **Error! Reference source not found.**(۲) نشان داده شده است.





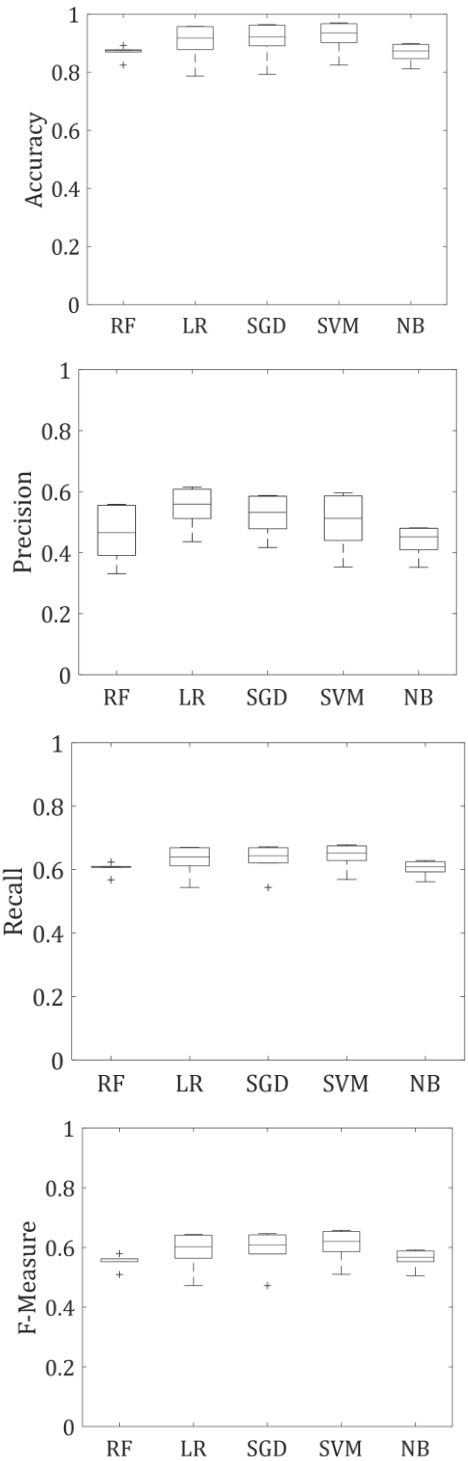
(شکل -۲). نتایج پارامترهای ارزیابی الگوریتم‌های یادگیری ماشین بر اساس تکنیک‌های چندواژه‌ای (دیجی‌کالا)

(Figure - 2). Machine learning algorithms result based on n-gram technics on Digikala comments

همان‌طور که در **Error! Reference source not found.**(۲) نشان داده شده است، مقدار دقت به‌دست‌آمده با تکنیک تک‌واژه‌ای در الگوریتم ماشین بردار پشتیبان بین ۲ تا ۱۳ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش تک‌واژه‌ای به دلیل این‌که به‌صورت کلمه به کلمه پردازش می‌شود درنتیجه نسبت به روش‌های دیگر نتایج بهتری را به دست می‌آورد. مقدار دقت به‌دست‌آمده با تکنیک دوواژه‌ای در الگوریتم ماشین بردار پشتیبان بین ۱ تا ۳ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش دوواژه‌ای به دلیل این‌که به‌صورت دوکلمه‌ای پردازش می‌شود درنتیجه نسبت به روش‌های دیگر به‌غیراز تک‌واژه‌ای نتایج بهتری را به دست می‌آورد. همچنین مقدار دقت به‌دست‌آمده با تکنیک سه‌واژه‌ای در الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی بین ۱ تا ۴ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. روش سه‌واژه‌ای به دلیل این‌که به‌صورت سه‌کلمه‌ای پردازش می‌شود و کلمات چند بار تکرار می‌شوند؛ بنابراین، بر احتمال سند اثر می‌گذارد درنتیجه نسبت به روش‌های دیگر دقت دسته‌بندی را کاهش می‌دهد. مقدار دقت به‌دست‌آمده با تکنیک‌های تک‌واژه‌ای و دوواژه‌ای در الگوریتم‌های ماشین بردار پشتیبان و کاهش گرادیان تصادفی بین ۱ تا ۹ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. روش ترکیبی تک‌واژه‌ای و دوواژه‌ای به دلیل وجود روش تک‌واژه‌ای و دوواژه‌ای درنتیجه نسبت به روش‌های دیگر نیز نتایج بهتری را به دست می‌آورند. مقدار دقت به‌دست‌آمده با تکنیک‌های دوواژه‌ای و سه‌واژه‌ای در الگوریتم ماشین بردار پشتیبان بین ۱ تا ۳ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش ترکیبی دوواژه‌ای و سه‌واژه‌ای به دلیل تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر نیز نتایج ضعیف‌تری را به دست آورد. درنهایت مقدار دقت به‌دست‌آمده با سه تکنیک در الگوریتم‌های ماشین بردار پشتیبان و کاهش گرادیان تصادفی و رگرسیون لجستیک بین ۶ تا ۹ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. روش ترکیبی تک‌واژه‌ای و دوواژه‌ای و سه‌واژه‌ای به دلیل تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر نیز نتایج ضعیف‌تر و تأثیر روش تک‌واژه‌ای و دوواژه‌ای نتایج بهتری را به دست آورند؛ بنابراین نتایج خوبی را به ارمغان می‌آورند.

به‌طور خلاصه مقدار دقت به‌دست‌آمده در الگوریتم‌های ماشین بردار پشتیبان در همه روش‌ها به‌غیراز روش سه‌واژه‌ای به میزان ۷ درصد و کاهش گرادیان تصادفی و رگرسیون لجستیک در همه روش‌ها به‌غیراز دوواژه‌ای و سه‌واژه‌ای و روش ترکیبی دوواژه‌ای و سه‌واژه‌ای به میزان ۵ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر در الگوریتم‌های مختلف نیز نتایج ضعیف‌تری را به دست‌آورند. از سوی دیگر صحت دقیق بودن نتیجه دسته‌بندی است بنابراین هر چه مقدار صحت بیشتر باشد عملکرد دسته‌بندی دقیق‌تر خواهد بود. همان‌طور که مشاهده می‌شود مقدار صحت به‌دست‌آمده در الگوریتم رگرسیون لجستیک در همه روش‌ها به‌غیراز روش دوواژه‌ای و سه‌واژه‌ای و ترکیب دوواژه‌ای و سه‌واژه‌ای به میزان ۲ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. همچنین بازیابی کامل بودن نتیجه دسته‌بندی است بنابراین نسبت به صحت هر چه مقدار بازیابی کمتر باشد عملکرد دسته‌بندی کامل‌تر خواهد بود. مقدار بازیابی به‌دست‌آمده روش سه‌واژه‌ای در همه الگوریتم‌های یادگیری ماشین به میزان ۱۱ درصد نتایج بهتری از مقدار حاصل‌شده در روش‌های دیگر را به دست می‌آورد. معیار فیشر متوسط صحت و بازیابی است. لازم است که سیستم را به سمت صحت یا بازیابی بهینه‌سازی کنیم که تأثیر

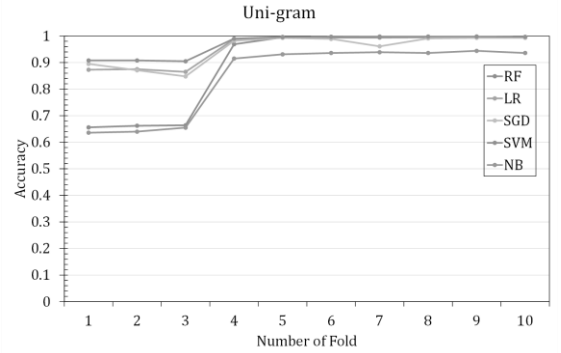
بیشتری بر نتیجه نهایی دارد. مقدار معیار فیشر به‌دست‌آمده در الگوریتم‌های ماشین بردار پشتیبان و کاهش گرادیان تصادفی و رگرسیون لجستیک در همه روش‌ها به‌غیراز روش دوواژه‌ای و سه‌واژه‌ای و ترکیب دوواژه‌ای و سه‌واژه‌ای به میزان ۱۰ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. برای درک بیشتر نمودار آماری الگوریتم‌های مختلف ازجمله میانگین، انحراف معیار و چارک‌ها در شکل - ۱) آمده است.

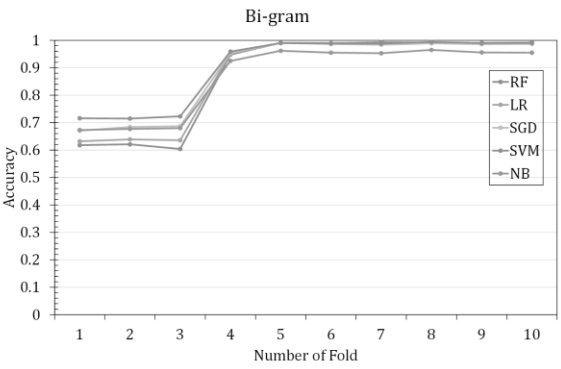


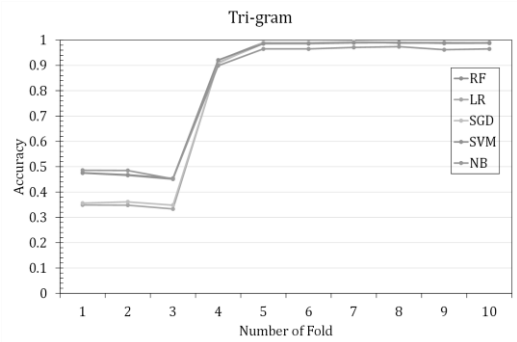
(شکل - ۱) .نمودار آماری الگوریتم‌های مختلف ازجمله میانگین، انحراف معیار و چارک‌ها (دیجی‌کالا)

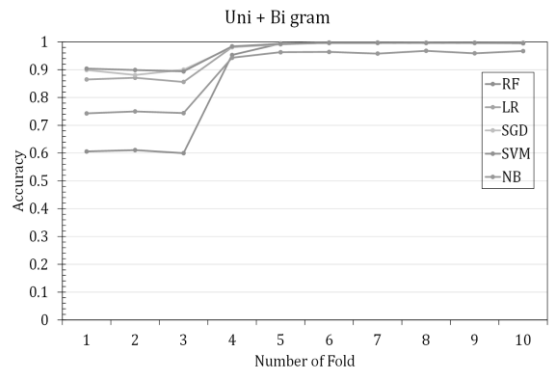
(Figure - 3). Statistical charts of different algorithms such as Average, SD, and Quarters (Digikala)

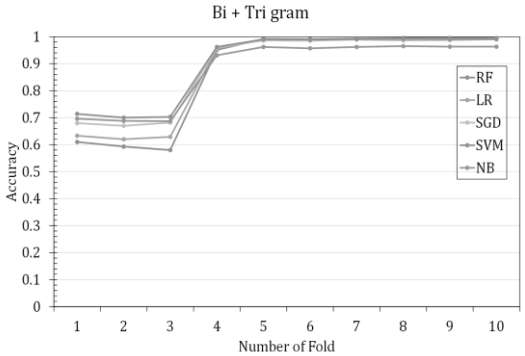
دقت به‌دست‌آمده پس از اعمال روش پیشنهادی و با استفاده از الگوریتم‌های مختلف یادگیری ماشین با آزمایش در ۱۰ دور متفاوت در **Error! Reference source not found.**(۴) نشان داده شده است. مشاهده می‌شود که مقدار دقت به‌دست‌آمده در الگوریتم‌های ماشین بردار پشتیبان به میزان ۲۵ درصد و کاهش گرادیان تصادفی به میزان ۲۴ درصد و رگرسیون لجستیک به میزان ۲۲ درصد از دور اول و الگوریتم‌های بیز ساده به میزان ۲۶ درصد و جنگل تصادفی به میزان ۳۰ درصد از دور چهارم به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش دوواژه‌ای پس از اعمال روش پیشنهادی و با استفاده از الگوریتم‌های مختلف یادگیری ماشین در همه الگوریتم‌های یادگیری ماشین از دور چهارم بین ۲۳ تا ۳۴ درصد به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش سه‌واژه‌ای پس از اعمال روش پیشنهادی از دور چهارم بین ۴۷ تا ۵۸ درصد و الگوریتم بیز ساده از دور پنجم به میزان ۷ درصد به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش ترکیبی تک‌واژه‌ای و دوواژه‌ای در الگوریتم‌های ماشین بردار پشتیبان به میزان ۳۰ درصد و کاهش گرادیان تصادفی به میزان ۲۹ درصد و رگرسیون لجستیک به میزان ۲۶ درصد از دور اول و الگوریتم‌های بیز ساده به میزان ۲۰ درصد و جنگل تصادفی به میزان ۳۵ درصد از دور چهارم به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش ترکیبی دوواژه‌ای و سه‌واژه‌ای در همه الگوریتم-های یادگیری ماشین از دور چهارم بین ۲۵ تا ۳۲ درصد به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش ترکیبی سه تکنیک در الگوریتم‌های ماشین بردار پشتیبان و کاهش گرادیان تصادفی و رگرسیون لجستیک از دور اول به میزان ۳۰ درصد و الگوریتم‌های بیز ساده به میزان ۲۱ درصد و جنگل تصادفی به میزان ۳۵ درصد از دور چهارم به سمت بالا بهبود یافته‌اند.

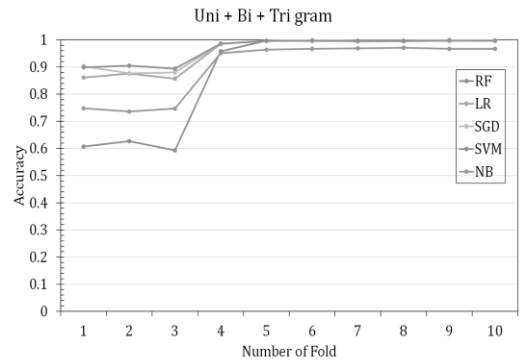








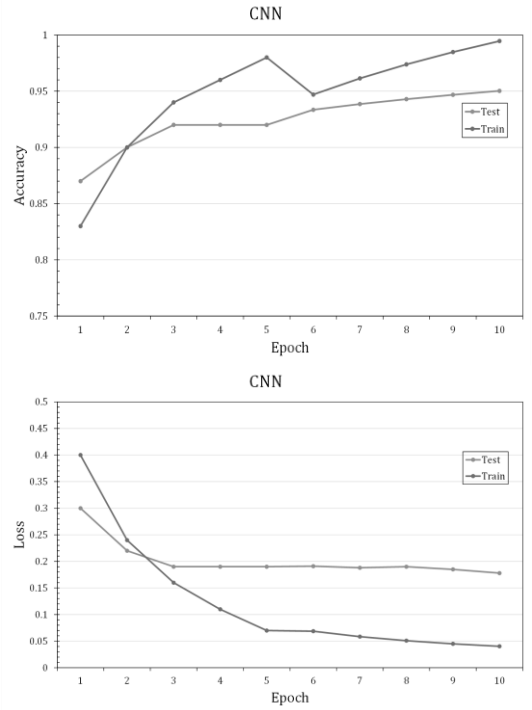




(شکل – ۴) نتایج دقت روش بر اساس دورهای مختلف اجرا با استفاده از الگوریتم‌های یادگیری ماشین (دیجی‌کالا)

(Figure - 4). Accuracy result of execution rounds using different n-gram technics (Digikala)

دقت و خطای به‌دست‌آمده در روش شبکه عصبی کانولوشن از روش‌های یادگیری عمیق پس از اعمال روش پیشنهادی در **Error! Reference source not found.** (۵) نشان داده شده است. می‌توان تحلیل کرد که مقدار دقت به‌دست‌آمده بین آموزش و آزمایش تا حدودی از بین رفته است و هرچند در دور سوم به میزان ۲ درصد افزایش دقت در داده‌های آموزش داریم ولی در داده‌های آزمایش ثابت شده که حالت بیش‌برازش پیدا کرده است و دقت از آموزش کمتر شده است. هم‌چنین درصد خطای روش شبکه عصبی کانولوشن به‌دست‌آمده در داده‌های آموزش در هر دوره بین ۴ تا ۲۵ درصد کمتر شده، ولی برای داده‌های آزمایش در دوره سوم با ۱۹ درصد خطا ثابت شده و بیشتر از داده‌های آموزش است. به‌طورکلی روش CNN بر روی داده‌های موردنظر و بعد از اعمال روش پیشنهادی نسبت به الگوریتم‌های جنگل تصادفی و بیز ساده با ۹۲ درصد دقت بهتر عمل کرده است.

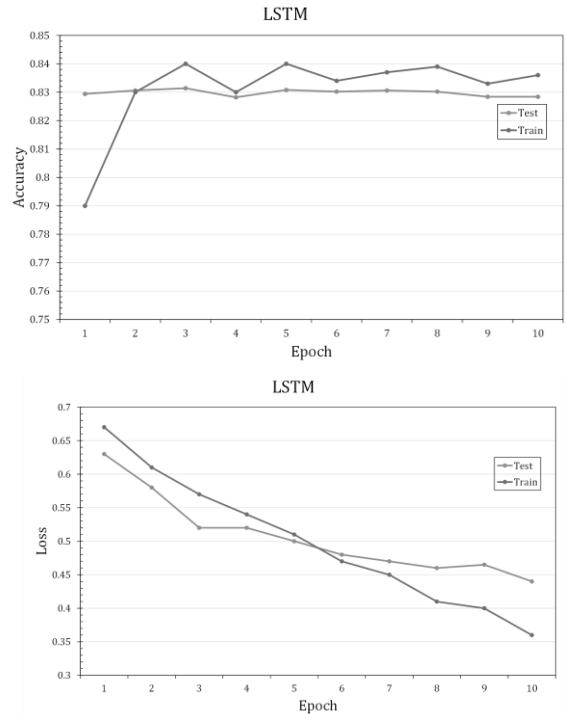


(شکل – ۵) . دقت و درصد خطای به‌دست‌آمده با روش شبکه عصبی کانولوشن (دیجی‌کالا)

(Figure - 5). Accuracy and loss using Convolution Neural Network (Digikala)

دقت و خطای به‌دست‌آمده در روش حافظه کوتاه‌مدت متوالی از روش‌های یادگیری عمیق پس از اعمال روش پیشنهادی در **Error! Reference source not found.**(۶) نشان داده شده است. مقدار دقت به‌دست‌آمده بین آموزش و آزمایش در همان دور اول هم در آموزش و هم در

آزمایش با مقدار ۸۳ درصد ثابت شده‌اند و حالت بیش‌برازش پیدا کرده‌اند. هم‌چنین درصد خطای روش حافظه کوتاه‌مدت متوالی در داده‌های آموزش و آزمایش در هر دوره بین ۲ تا ۶ درصد کمتر شده است و در دوره‌های آخر در داده‌های آزمایش درصد خطا پایین‌تر نیز آمده است. به‌طورکلی روش LSTM بر روی داده‌های موردنظر و بعد از اعمال روش پیشنهادی نسبت به الگوریتم‌های یادگیری ماشین و CNN با ۸۳ درصد دقت به دلیل این‌که وابستگی جملات را لحاظ نمی‌کند ضعیف‌تر عمل کرده است.

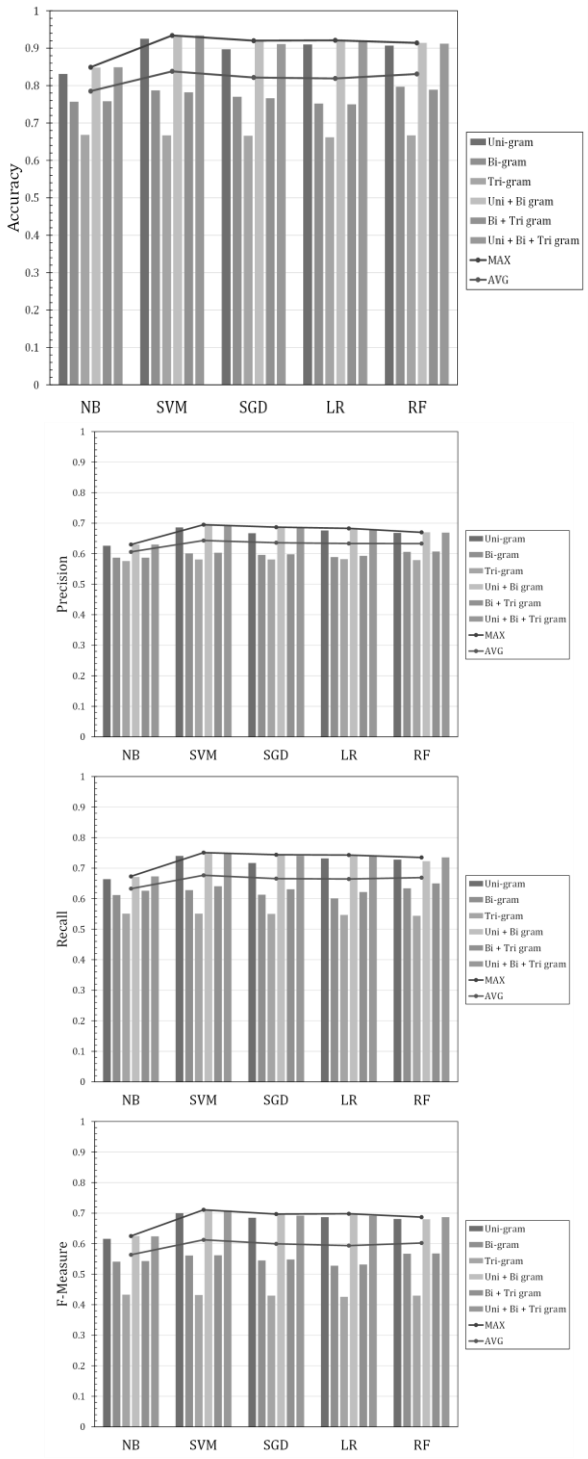


(شکل – ۶). دقت و خطای به‌دست‌آمده در روش حافظه کوتاه‌مدت متوالی (دیجی‌کالا)

(Figure - 6). Accuracy and loss using LSTM Network (Digikala)

۴-۳- تأثیر استفاده از فرهنگ‌نامه و متوازن‌سازی بر داده‌های سنتی‌پرس

در این قسمت نتایج پارامترهای ارزیابی نظرات تلفن همراه از سایت سنتی‌پرس با استفاده از الگوریتم‌های یادگیری ماشین توسط تکنیک‌های تک‌واژه‌ای، دوواژه‌ای، سه‌واژه‌ای، ترکیب تک‌واژه‌ای و دوواژه‌ای، ترکیب دوواژه‌ای و سه‌واژه‌ای و ترکیب هر سه تکنیک بررسی شده است. با استفاده از ماتریس درهم‌ریختگی، پارامترهای مختلف ارزیابی مانند دقت، صحت، بازیابی و معیار فیشر به‌دست‌آمده پس از اعمال روش پیشنهادی محاسبه شده است. دسته‌بندی نتایج بر اساس انواع تکنیک‌های چندواژه‌ای و با استفاده از الگوریتم‌های مختلف یادگیری ماشین در **Error! Reference source not found.**(۷) نشان داده شده است.



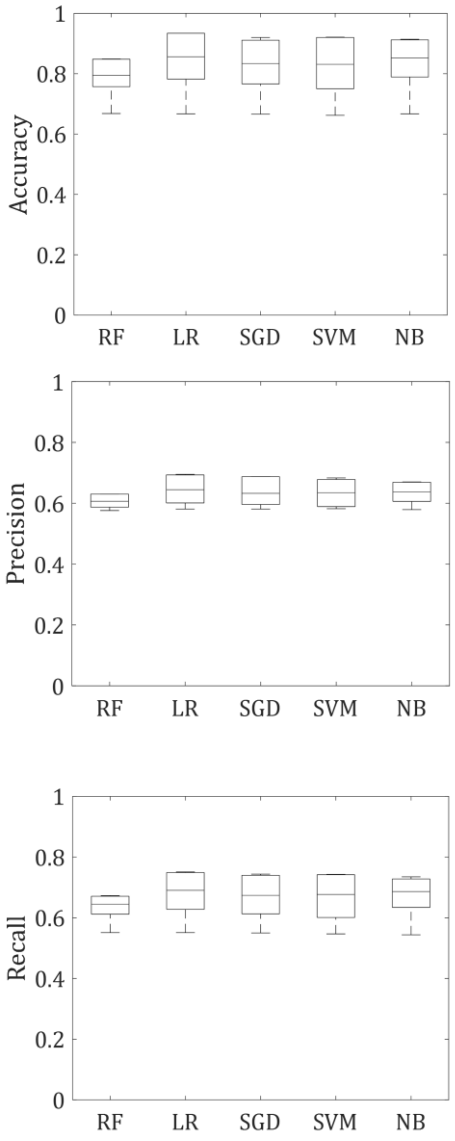
(شکل – ۷). نتایج پارامترهای ارزیابی الگوریتم‌های یادگیری ماشین بر اساس تکنیک‌های چندواژه‌ای (سنتی‌پرس)

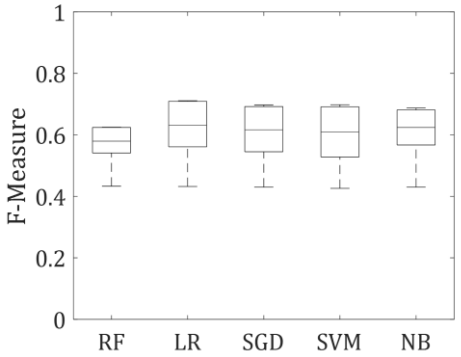
(Figure - 7). Machine learning algorithms result based on n-gram technics on Senti-Pers comments

همان‌طور که در **Error! Reference source not found.**(۲) نشان داده شده است، مقدار دقت به‌دست‌آمده با تکنیک تک‌واژه‌ای در الگوریتم ماشین بردار پشتیبان بین ۱ تا ۹ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش تک‌واژه‌ای به دلیل این‌که به‌صورت

کلمه به کلمه پردازش می‌شود درنتیجه نسبت به روش‌های دیگر نتایج بهتری را به دست می‌آورد. مقدار دقت به‌دست‌آمده با تکنیک دوواژه‌ای در الگوریتم جنگل تصادفی بین ۱ تا ۴ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش دوواژه‌ای به دلیل این‌که به‌صورت دوکلمه‌ای پردازش می‌شود درنتیجه نسبت به روش‌های دیگر به‌غیراز تک‌واژه‌ای نتایج بهتری را به دست می‌آورد. همچنین مقدار دقت به‌دست‌آمده با تکنیک سه‌واژه‌ای در همه الگوریتم‌های یادگیری ماشین به میزان ۶۶ درصد یکسان هستند. روش سه‌واژه‌ای به دلیل این‌که به‌صورت سه‌کلمه‌ای پردازش می‌شود و کلمات چند بار تکرار می‌شوند؛ بنابراین، بر احتمال سند اثر می‌گذارد درنتیجه نسبت به روش‌های دیگر دقت دسته‌بندی را کاهش می‌دهد. مقدار دقت به‌دست‌آمده با تکنیک‌های تک‌واژه‌ای و دوواژه‌ای در الگوریتم ماشین بردار پشتیبان بین ۱ تا ۹ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. روش ترکیبی تک‌واژه‌ای و دوواژه‌ای به دلیل وجود روش تک‌واژه‌ای و دوواژه‌ای درنتیجه نسبت به روش‌های دیگر نیز نتایج بهتری را به دست می‌آورند. مقدار دقت به‌دست‌آمده با تکنیک‌های دوواژه‌ای و سه‌واژه‌ای در الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی بین ۲ تا ۳ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر است. روش ترکیبی دوواژه‌ای و سه‌واژه‌ای به دلیل تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر نیز نتایج ضعیف‌تری را به دست آورد. دقت به‌دست‌آمده با سه تکنیک در الگوریتم ماشین بردار پشتیبان بین ۱ تا ۹ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. روش ترکیبی تک‌واژه‌ای و دوواژه‌ای و سه‌واژه‌ای به دلیل تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر نیز نتایج ضعیف‌تر و تأثیر روش تک‌واژه‌ای و دوواژه‌ای نتایج بهتری را به دست آورند؛ بنابراین نتایج خوبی را به ارمغان می‌آورند.

به‌طور خلاصه دقت به‌دست‌آمده پس از اعمال روش پیشنهادی و استفاده از الگوریتم‌های مختلف یادگیری ماشین مقدار دقت به‌دست‌آمده در الگوریتم‌های ماشین بردار پشتیبان در همه روش‌های N-Gram به‌غیراز روش سه‌واژه‌ای به میزان ۴ درصد و رگرسیون لجستیک و جنگل تصادفی در همه روش‌ها به‌غیراز دوواژه‌ای و سه‌واژه‌ای و روش ترکیبی دوواژه‌ای و سه‌واژه‌ای به میزان ۲ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. تأثیر روش سه‌واژه‌ای باعث می‌شود که مقدار دقت نسبت به روش‌های دیگر در الگوریتم‌های مختلف نیز نتایج ضعیف‌تری را به دست آورند. مقدار صحت به‌دست‌آمده در الگوریتم ماشین بردار پشتیبان در همه روش‌ها به‌غیراز روش سه‌واژه‌ای به میزان ۶ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر در روش‌های دیگر هستند. همچنین مقدار بازیابی به‌دست‌آمده به‌غیراز روش سه‌واژه‌ای در همه الگوریتم‌های یادگیری ماشین به میزان ۱۹ درصد نتایج بهتری را به دست می‌آورد. معیار فیشر به‌دست‌آمده در الگوریتم‌های ماشین بردار پشتیبان در همه روش‌ها به‌غیراز روش سه‌واژه‌ای به میزان ۹ درصد و کاهش گرادیان تصادفی و رگرسیون لجستیک و جنگل تصادفی در همه روش‌ها به‌غیراز روش سه‌واژه‌ای به میزان ۷ درصد بهتر از مقدار حاصل‌شده در الگوریتم‌های دیگر هستند. برای درک بیشتر نمودار آماری الگوریتم‌های مختلف ازجمله میانگین، انحراف معیار و چارک‌ها در **Error! Reference source not found.**(۸) آمده است.

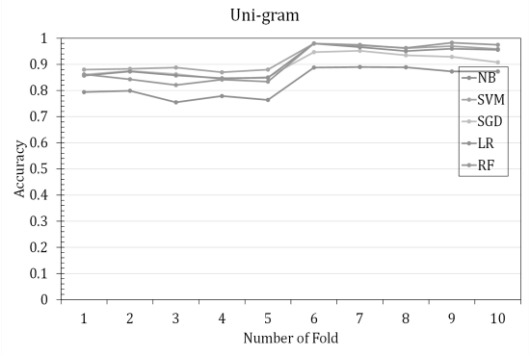


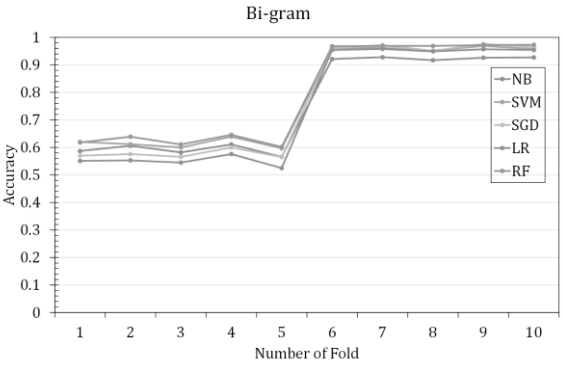


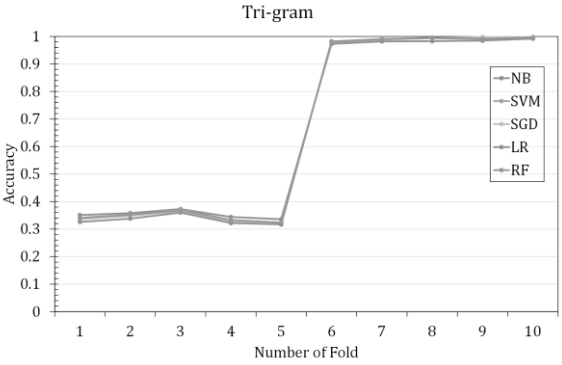
(شکل – ۸). نمودار آماری الگوریتم‌های مختلف ازجمله میانگین، انحراف معیار و چارک‌ها (سنتی‌پرس)

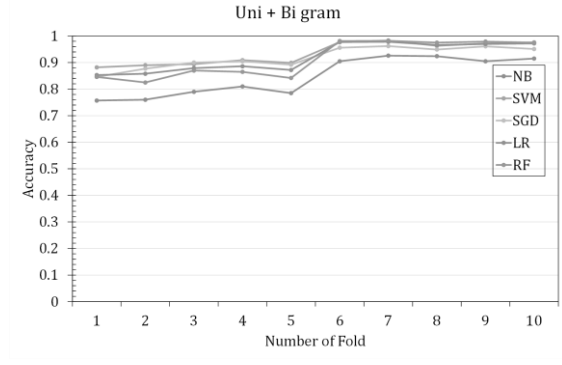
(Figure - 8). Statistical charts of different algorithms such as Average, SD, and Quarters (Senti-Pers)

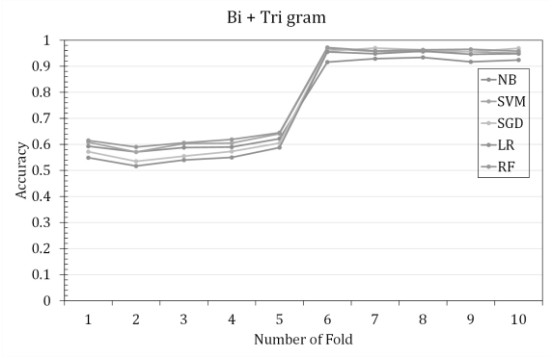
دقت به‌دست‌آمده پس از اعمال روش پیشنهادی و با استفاده از الگوریتم‌های مختلف یادگیری ماشین با آزمایش در ۱۰ دور متفاوت در **Error! Reference source not found.**(۹) نشان داده شده است. مشاهده می‌شود که مقدار دقت به‌دست‌آمده در الگوریتم‌های بیز ساده به میزان ۱۲ درصد و ماشین بردار پشتیبان به میزان ۱۰ درصد و کاهش گرادیان تصادفی به میزان ۹ درصد و رگرسیون لجستیک به میزان ۱۳ درصد و جنگل تصادفی به میزان ۱۵ درصد از دور ششم به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش دوواژه‌ای پس از اعمال روش پیشنهادی در همه الگوریتم‌های یادگیری ماشین از دور ششم بین ۳۶ تا ۴۳ درصد به سمت بالا بهبود یافته‌اند. همچنین دقت به‌دست‌آمده در روش سه‌واژه‌ای در همه الگوریتم‌های یادگیری ماشین از دور ششم بین ۶۴ تا ۶۷ درصد به سمت بالا بهبود یافته‌اند. این مقدار در روش ترکیبی تک‌واژه‌ای و دوواژه‌ای در الگوریتم‌های بیز ساده به میزان ۱۲ درصد و ماشین بردار پشتیبان به میزان ۸ درصد و کاهش گرادیان تصادفی به میزان ۶ درصد و رگرسیون لجستیک به میزان ۱۰ درصد و جنگل تصادفی به میزان ۱۴ درصد از دور ششم به سمت بالا بهبود یافته‌اند. دقت به‌دست‌آمده در روش ترکیبی دوواژه‌ای و سه‌واژه‌ای پس از اعمال روش پیشنهادی در همه الگوریتم‌های یادگیری ماشین از دور ششم بین ۲۹ تا ۳۷ درصد به سمت بالا بهبود یافته‌اند. درنهایت دقت به‌دست‌آمده در روش ترکیبی سه تکنیک در الگوریتم‌های ماشین بردار پشتیبان به میزان ۹ درصد و کاهش گرادیان تصادفی به میزان ۶ درصد از دور پنجم و بیز ساده به میزان ۱۲ درصد و رگرسیون لجستیک به میزان ۱۱ درصد و جنگل تصادفی به میزان ۱۳ درصد از دور ششم به سمت بالا بهبود یافته‌اند.

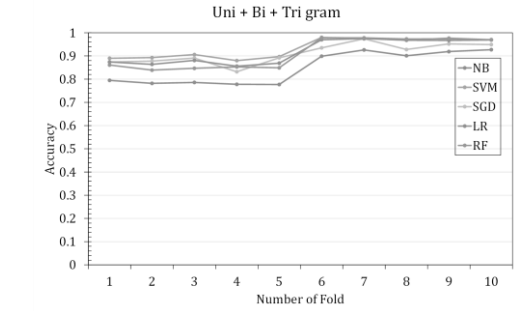










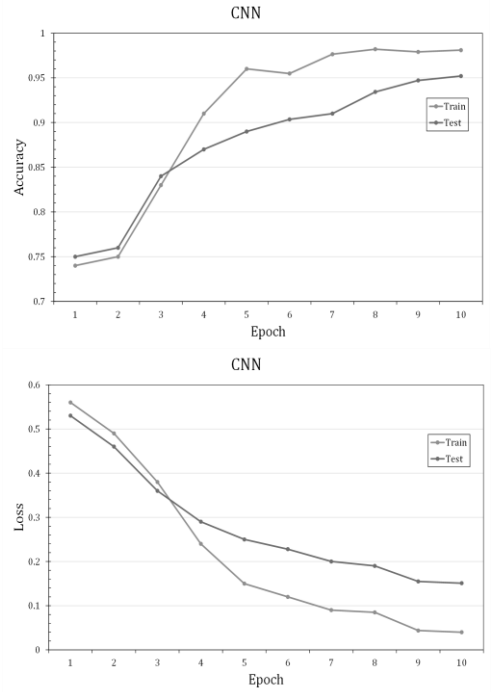


(شکل – ۹). نتایج دقت روش بر اساس دورهای مختلف اجرا با استفاده از الگوریتم‌های یادگیری ماشین (سنتی‌پرس)

(Figure - 9). Accuracy result of execution rounds using different n-gram technics (Senti-Pers)

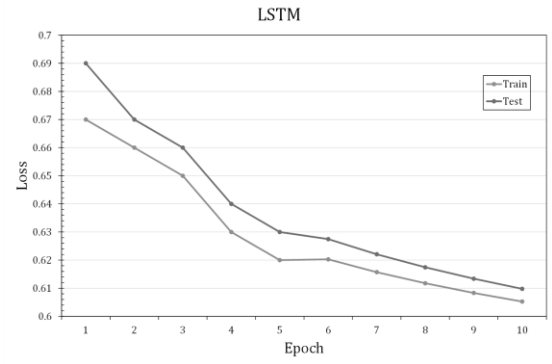
دقت و خطای به‌دست‌آمده در روش شبکه عصبی کانولوشن از روش‌های یادگیری عمیق پس از اعمال روش پیشنهادی در **Error! Reference source not found.**(۱۰) نشان داده شده است. همان‌طور که نشان داده شده است، می‌توان تحلیل کرد که مقدار دقت به‌دست‌آمده بین آموزش و آزمایش تا حدودی از بین رفته است و هرچند در دور پنجم به میزان ۵ درصد افزایش دقت در داده‌های آموزش داریم ولی در داده‌های آزمایش این مقدار از بین رفته است و بیش‌برازش نمایان شده است. هم‌چنین درصد خطای روش شبکه عصبی کانولوشن در داده‌های آموزش در هر دوره بین ۷ تا ۱۴ درصد کمتر شده است ولی برای داده‌های آزمایش در هر دوره بین ۴ تا ۱۰ درصد خطا کمتر شده است و در دوره آخر با میزان ۲۵ بیشتر از داده‌های آموزش است. به‌طورکلی روش CNN بر روی داده‌های موردنظر و بعد از اعمال روش پیشنهادی نسبت به الگوریتم‌های رگرسیون لجستیک و جنگل تصادفی و بیز ساده با ۸۹ درصد دقت بهتر عمل کرده است.

دقت و خطای به‌دست‌آمده در روش حافظه کوتاه‌مدت متوالی از روش‌های یادگیری عمیق پس از اعمال روش پیشنهادی در **Error! Reference source not found.**(۱۱) نشان داده شده است. مشاهده می‌شود که مقدار دقت به‌دست‌آمده در داده‌های آموزش بین ۲ تا ۱۲ تغییر یافته است ولی در داده‌های آزمایش در دوره سوم با مقدار ۷۵ درصد ثابت شده است و حالت بیش‌برازش پیدا کرده‌اند. هم‌چنین درصد خطای روش حافظه کوتاه‌مدت متوالی در داده‌های آموزش و آزمایش در هر دوره بین ۱ تا ۲ درصد کمتر شده است و در دوره آخر در داده‌های آزمایش درصد خطا با مقدار ۶۲ درصد پایین‌تر از آموزش نیز آمده است. به‌طورکلی روش LSTM بر روی داده‌های موردنظر و بعد از اعمال روش پیشنهادی نسبت به الگوریتم‌های یادگیری ماشین و CNN با ۷۵ درصد دقت به دلیل این‌که وابستگی جملات را لحاظ نمی‌کند ضعیف‌تر عمل کرده است.



(شکل – ۱۰). دقت و درصد خطای به‌دست‌آمده با روش شبکه عصبی کانولوشن (سنتی‌پرس)

(Figure - 10). Accuracy and loss using Convolution Neural Network (Senti-Pers)



(شکل – ۱۱). دقت و خطای به‌دست‌آمده در روش حافظه کوتاه‌مدت متوالی (سنتی‌پرس)

(Figure - 11). Accuracy and loss using LSTM (Senti-Pers)

۴-۴- مقایسه دقت رویکرد پیشنهادی با کارهای دیگران

نتیجه مقایسه‌ای دقت رویکرد پیشنهادی با کارهای مختلف و استفاده از داده‌های فارسی تلفن همراه از سایت دیجی‌کالا و داده‌های سنتی‌پرس از سایت پیکره‌گان در جدول (۲) آمده است. این اطلاعات بر اساس روش‌های مختلف چندواژه‌ای و الگوریتم‌های مختلف یادگیری ماشین دسته‌بندی شده است. اگرچه نویسندگان واقف‌اند که در مورد مقایسه با کارهای گذشته شاید این مقایسه معنادار نباشد چراکه مجموعه‌داده‌ها متفاوت است ولی می‌توان با یک نگاه کلی‌نگر تا حدودی نتیجه نهایی را تحلیل نمود. همان‌طور که مشاهده می‌شود، می‌توان تحلیل کرد که مقدار دقت به‌دست‌آمده با استفاده از روش پیشنهادی روی داده‌های دیجی‌کالا و سنتی‌پرس نسبت به روش قبل از اعمال فرهنگ‌نامه و متوازن‌سازی داده‌ها و کارهای دیگران در همه تکنیک‌های چندواژه‌ای و تمام الگوریتم‌های یادگیری ماشین بهتر بوده است. هم‌چنین می‌توان دریافت که دقت به‌دست‌آمده از روش اعمال پیشنهادی روی داده‌های دیجی‌کالا در الگوریتم‌های بیز ساده بین ۱۰ تا ۳۴ درصد و ماشین بردار پشتیبان بین ۵ تا ۲۴ درصد و کاهش گرادیان تصادفی بین ۷ تا ۳۸ درصد و رگرسیون لجستیک بین ۵ تا ۳۸ درصد و جنگل تصادفی بین ۴ تا ۲۲ درصد در همه روش‌ها نتایج بهتری را به دست آورده‌اند. همچنین در روش‌های یادگیری عمیق روش CNN با اعمال روش پیشنهادی روی داده‌های دیجی‌کالا به میزان ۴ درصد بهبود یافته است ولی در روش LSTM هیچ تغییری نیافته است. هم‌چنین دقت به‌دست‌آمده از اعمال روش پیشنهادی روی داده‌های سنتی‌پرس در الگوریتم‌های بیز ساده بین ۱۲ تا ۴۶ درصد و ماشین بردار پشتیبان بین ۵ تا ۴۶ درصد و کاهش گرادیان تصادفی بین ۵ تا ۳۵ درصد و رگرسیون لجستیک بین ۶ تا ۴۶ درصد و جنگل تصادفی بین ۴ تا ۴۶ درصد در همه روش‌ها نتایج بهتری را به دست آورده‌اند.

(جدول - ۲) نتیجه مقایسه دقت با ادبیات مختلف و داده‌های فارسی تلفن همراه از سایت دیجی‌کالا و داده‌های سنتی‌پرس
(Table - 2). Accuracy comparison with related works

		قبل از اعمال روش جدید			رویکرد جدید	
		دیجی‌کالا	سنتی‌پرس	دیجی‌کالا	سنتی‌پرس	
Pang	تک‌واژه	۷۹.۵	۸۱	۸۴.۷	۶۸.۴	۸۳.۱
Salvetti	دوواژه	–	۷۷.۳	۸۷	۵۴.۴	۷۵.۷
Beineke	سه‌واژه	–	–	۸۱.۲	۲۰.۷	۶۶.۸
Mullen, Collier	تک و دوواژه	–	–	۸۹.۶	۷۲.۱	۸۴.۸
Matsumoto	دو و سه‌واژه	–	–	۸۷.۷	۵۳.۴	۷۵.۸
بیز ساده	هر سه تکنیک	–	–	۸۹.۹	۷۱.۵	۸۴.۹
	تک‌واژه	۷۲.۹	–	۹۷	۸۷.۳	۹۲.۵
	دوواژه	۷۷.۱	–	۹۰.۵	۵۹.۴	۷۸.۷
	سه‌واژه	–	–	۸۲.۵	۲۰.۵	۶۶.۷
	تک و دوواژه	–	–	۹۶.۶	۸۸.۳	۹۳.۴
ماشین بردار پشتیبان	دو و سه‌واژه	–	–	۹۰.۲	۵۹.۴	۷۸.۲
	هر سه تکنیک	–	–	۹۶.۷	۸۸.۵	۹۳.۴
	تک‌واژه	–	–	۹۵.۱	۸۳.۵	۸۹.۷
	دوواژه	–	–	۸۹.۲	۵۷.۳	۷۷
	سه‌واژه	–	–	۷۹.۳	۳۱.۹	۶۶.۶
	تک و دوواژه	–	–	۹۶.۴	۸۶.۵	۹۲
کاهش گرادیان تصادفی	دو و سه‌واژه	–	–	۸۹.۱	۵۷.۱	۷۶.۶
	هر سه تکنیک	–	–	۹۶.۲	۸۶.۵	۹۱.۱
	تک‌واژه	–	–	۹۵.۸	۸۳.۳	۹۱
	دوواژه	–	–	۸۸	۵۵.۲	۷۵.۲
	سه‌واژه	–	–	۷۸.۷	۲۰	۶۶.۲
	تک و دوواژه	–	–	۹۵.۵	۸۵.۵	۹۲.۱
رگرسیون لجستیک	دو و سه‌واژه	–	–	۸۷.۸	۵۵	۷۵
	هر سه تکنیک	–	–	۹۵.۷	۸۶	۹۲
	تک‌واژه	–	–	۸۹.۲	۸۶.۹	۹۰.۷
	دوواژه	–	–	۸۷.۴	۶۰.۱	۷۹.۷
	سه‌واژه	–	–	۸۲.۵	۲۰.۶	۶۶.۷
	تک و دوواژه	–	–	۸۷.۵	۸۶	۹۱.۴
جنگل تصادفی	دو و سه‌واژه	–	–	۸۷	۶۰.۱	۷۸.۹
	هر سه تکنیک	–	–	۸۷.۷	۸۶.۷	۹۱.۲
	شبکه عصبی کانولوشن	–	–	۹۲	۸۹	۸۹
یادگیری عمیق	حافظه کوتاه‌مدت متوالی (LSTM)	–	–	۸۳	۷۵	۷۵

۵- نتیجه‌گیری و کارهای آینده

این پژوهش تلاشی برای دسته‌بندی نظرات فارسی تلفن همراه از سایت دیجی‌کالا و داده‌های سنتی‌پرس از سایت پیکره‌گان با استفاده از الگوریتم‌های مختلف یادگیری ماشین تحت نظارت مانند بیز ساده (NB)، کاهش گرادیان تصادفی(SGD)، ماشین بردار پشتیبان(SVM) ، رگرسیون لجستیک (LR) و جنگل تصادفی (RF) و از الگوریتم‌های یادگیری عمیق مانند شبکه عصبی کانولوشن (CNN) و حافظه کوتاه‌مدت متوالی (LSTM) است. این الگوریتم‌ها با

استفاده از روش‌های چندواژه‌ای در مجموعه‌داده‌های تلفن همراه به کار می‌روند. مشاهده می‌شود که با افزایش مقدار واژه‌ها، دقت دسته‌بندی کاهش می‌یابد، یعنی برای تک-واژه‌ای و دوواژه‌ای، نتیجه استفاده از الگوریتم‌ها بهتر از سه‌واژه‌ای است.؛ نابراین الگوریتم‌های بیز ساده بین ۱۰ تا ۳۴ درصد و ماشین بردار پشتیبان بین ۵ تا ۲۴ درصد و کاهش گرادیان تصادفی بین ۷ تا ۳۸ درصد و رگرسیون لجستیک بین ۵ تا ۳۸ درصد و جنگل تصادفی بین ۴ تا ۲۲ درصد در همه روش‌های چندواژه‌ای و روش شبکه عصبی کانولوشن به میزان ۴ درصد روی داده‌های دیجی-کالا و داده‌های سنتی‌پرس در الگوریتم‌های بیز ساده بین ۱۲ تا ۴۶ درصد و ماشین بردار پشتیبان بین ۵ تا ۴۶ درصد و کاهش گرادیان تصادفی بین ۵ تا ۳۵ درصد و رگرسیون لجستیک بین ۶ تا ۴۶ درصد و جنگل تصادفی بین ۴ تا ۴۶ درصد در همه روش‌های چندواژه‌ای عملکرد بهتری نسبت به الگوریتم‌های دیگر داشتند. از روش‌های نمونه‌گیری بیش‌ازحد برای متوازن‌سازی داده‌ها و روش امتیازدهی پیشنهادی برای بهبود تعیین قطبیت و از تکنیک‌های بردارسازی برای تبدیل متن به ماتریس اعداد استفاده شده است و نیز به‌منظور به دست آوردن مقدار دقت، صحت، بازیابی و معیار فیشر در روش بهبودیافته، از تکنیک‌های یادگیری ماشین و یادگیری عمیق استفاده می‌شود. فرایند نظرکاوی در زبان فارسی مشکلاتی از قبیل عامیانه بودن کلمات و رعایت نکردن فاصله بین کلمات و رعایت نکردن قاعده ساختار جمله و استفاده از شکلک‌های تصویری را دارد که می‌توان این مشکلات را در کارهای آینده برای بهبود دسته‌بندی تحلیل احساسات رفع کرد.

6- Refrence ۶- مراجع

[1] Feldman, Ronen. "Techniques and applications for sentiment analysis." Communications of the ACM 56.4 (2013): 82-89.

[2] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." 2014 Seventh International Conference on Contemporary Computing (IC3). IEEE, 2014.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.

[4] Domingues, InĹes, et al. "Evaluation of oversampling data balancing techniques in the context of ordinal classification." 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018.

[5] Aung, Khin Zezawar, and Nyein Nyein Myo. "Sentiment analysis of students' comment using lexicon-based approach." 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS). IEEE, 2017.

[6] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "Unsupervised learning." The elements of statistical learning. Springer, New York, NY, 2009. 485-585.

[7] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).

[8] Salvetti, Franco, Stephen Lewis, and Christoph Reichenbach. "Automatic opinion polarity classification of movie reviews." Colorado research in linguistics 17 (2004).

[9] Beineke, Philip, Trevor Hastie, and Shivakumar Vaithyanathan. "The sentimental factor: Improving review classification via human-provided information." Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04). 2004.

[10] Mullen, Tony, and Nigel Collier. "Sentiment analysis using support vector machines with diverse information sources." Proceedings of the 2004 conference on empirical methods in natural language processing. 2004.

[11] Dave, Kushal, Steve Lawrence, and David M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews." Proceedings of the 12th international conference on World Wide Web. 2003.

[12] Matsumoto, Shotaro, Hiroya Takamura, and Manabu Okumura. "Sentiment classification using word sub-sequences and dependency sub-trees." Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, Heidelberg, 2005.

[13] Zhang, Dongwen, et al. "Chinese comments sentiment classification based on word2vec and SVMperf." Expert Systems with Applications 42.4 (2015): 1857-1863.

[14] Liu, Shuhua Monica, and Jiun-Hung Chen. "A multi-label Classification based approach for sentiment classification." Expert Systems with Applications 42.3 (2015): 1083-1093.

[15] Luo, Banghui, Jianping Zeng, and Jiangjiao Duan. "Emotion space model for classifying opinions in stock message board." Expert Systems with Applications 44 (2016): 138-146.

[16] Niu, Teng, et al. "Sentiment analysis on multi-view social data." International Conference on Multimedia Modeling. Springer, Cham, 2016.

[17] Li, Caiqiang, and Junming Ma. "Research on online education teacher evaluation model based on opinion mining." 2012 National Conference on Information Technology and Computer Science. Atlantis Press, 2012.

[18] Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro. "Sentiment analysis in Facebook and its application to e-learning." Computers in human behavior 31 (2014): 527-541.

[19] Pong-Inwong, Chakrit, and Wararat Songpan Rungworawut. "Teaching senti-lexicon for automated sentiment polarity definition in teaching evaluation." 2014 10th International Conference on Semantics, Knowledge and Grids. IEEE, 2014.

[20] Wang, Yili, and Hee Yong Youn. "Feature Weighting Based on Inter-Category and Intra-Category Strength for Twitter Sentiment Analysis." Applied Sciences 9.1 (2019): 92.

[21] Mnsefi, Gharizadeh, Sefidsangi, "An overview of opinion mining", The first specialized conference on intelligent computer systems and their applications, Tehran, 2011.

[22] Noeei, Jalali, Ghaemi, "Opinion mining: An overview of the work done", National Conference on Application of Intelligent Systems (soft computing) in Science and Technology, Quchan, 2013.

[23] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining." Lrec. Vol. 10. No. 2010. 2010.

[24] Ngoc, Phan Trong, and Myungsik Yoo. "The lexicon-based sentiment analysis for fan page ranking in Facebook." The International Conference on Information Networking 2014 (ICOIN2014). IEEE, 2014.

[25] Tang, Huifeng, Songbo Tan, and Xueqi Cheng. "A survey on sentiment detection of reviews." Expert Systems with Applications 36.7 (2009): 10760-10773.

[26] Garreta, Raul, and Guillermo Moncecchi. Learning scikit-learn: machine learning in python. Packt Publishing Ltd, 2013.

[27] McCallum, Andrew, and Kamal Nigam. "A comparison of event models for naive bayes text classification." AAAI-98 workshop on learning for text categorization. Vol. 752. No. 1. 1998.

[28] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1396-1400.

[29] Bottou, Léon. "Stochastic gradient descent tricks." Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012. 421-436.

[30] Walker, Strother H., and David B. Duncan. "Estimation of the probability of an event as a function of several independent variables." Biometrika 54.1-2 (1967): 167-179.

[31] Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[32] Yuan, Yufei, and Michael J. Shaw. "Induction of fuzzy decision trees." Fuzzy Sets and systems 69.2 (1995): 125-139.

[33] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[34] Hochreiter, Sepp. "JA1 4 rgen Schmidhuber (1997). "Long Short-Term Memory"." Neural Computation 9.8.

[35] Tripathy, Abinash, Ankit Agrawal, and Santanu Kumar Rath. "Classification of sentiment reviews using n-gram machine learning approach." Expert Systems with Applications 57 (2016): 117-126.

[36] Mouthami, K., K. Nirmala Devi, and V. Murali Bhaskaran. "Sentiment analysis and classification based on textual reviews." 2013 international conference on Information communication and embedded systems (ICICES). IEEE, 2013.

[37] Zhang, H. "The Optimality of Naive Bayes," In Proc. Seventeenth Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2004, vol. 1, no. 2, pp. 1-6. 2004.

[38] Bishop, Christopher M. Pattern recognition and machine learning. springer, 2006.

[39] Bottou, Léon. "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT'2010. Physica-Verlag HD, 2010. 177-186.

[40] Menard, Scott. Applied logistic regression analysis. Vol. 106. Sage, 2002.

[41] Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[42] Ketkar, Nikhil, and Eder Santana. Deep Learning with Python. Vol. 1. Berkeley, CA: Apress, 2017.

[43]

مهدیه واحدی پور کارشناس ارشد مهندسی رایانه گرایش نرم افزار از دانشگاه صنعتی قم است. زمینه های پژوهشی وی هوش مصنوعی، داده کاوی و پردازش متن است.

نشانی رایانامه ایشان عبارت است از:

vahedipoor.m@qut.ac.ir

محبوبه شمسی دکترای خود را در رشته مهندسی رایانه گرایش نرم‌افزار در سال ۱۳۹۰ دریافت کرده است. ایشان هم اکنون استادیار بخش

مهندسی رایانه دانشکده برق و رایانه در دانشگاه صنعتی قم است. نامبرده تاکنون مقالات علمی بسیاری را در مجلات و کنفرانس های معتبر خارجی و داخلی به چاپ رسانده است. زمینه های پژوهشی وی پردازش تصویر، یادگیری ماشین، یادگیری عمیق و پایگاه داده‌ها است.
نشانی رایانامه ایشان عبارت است از:
shamsi@qut.ac.ir

عبدالرضا رسولی کناری دکترای خود را در رشته مهندسی رایانه گرایش امنیت داده در سال ۱۳۹۰ دریافت کرده است. ایشان هم اکنون استادیار بخش مهندسی رایانه دانشکده برق و رایانه در دانشگاه صنعتی قم است. نامبرده تاکنون مقالات علمی بسیاری را در مجلات و کنفرانس های معتبر خارجی و داخلی به چاپ رسانده است. زمینه های پژوهشی وی امنیت و رمزنگاری، یادگیری ماشینُ یادگیری عمیق و کلان داده‌ها است.
نشانی رایانامه ایشان عبارت است از:
rasouli@qut.ac.ir