دادگان پرسش و پاسخ زبان فارسی

جواد فروتن راد*[۱]، مریم حور علی [۲]، محمدعلی کیوان راد [۳]
[۱]کارشناس ارشد هوش مصنوعی، دانشگاه صنعتی مالک‌اشتر
[۲ و ۳] استادیار هوش مصنوعی، مجتمع برق و کامپیوتر، دانشگاه صنعتی مالک‌اشتر

چکیده:

پاسخ سریع و دقیق به سؤالات مطرح‌شده به زبان طبیعی یکی از اهداف مهم در توسعه سامانه‌های پرسش‌وپاسخ است که در آن رایانه یک متن و سؤال را درک و پاسخ دقیق را برای کاربر ارائه می‌کند. با اینکه پیشرفت‌های زیادی در این حوزه صورت‌گرفته‌است، اما همچنان جزء مسائلی است که نیاز به ارتقا، به‌خصوص برای زبان‌های غیر انگلیسی مثل زبان فارسی است. در این مقاله دادگان پرسش‌وپاسخ زبان فارسی (FarsiQuAD) [۱] که توسط انسان از مقالات ویکی‌پدیای فارسی تهیه شده، در دو نسخه منتشر شده‌است. نسخه یک شامل ۱۰۰۰۰+ پرسش‌وپاسخ و نسخه دوم این مجموعه شامل بیش از ۱۴۵۰۰۰+ جفت پرسش‌وپاسخ است. این دادگان قابلیت تجمیع با نسخه انگلیسی SQuAD و سایر دادگان زبان‌های دیگر را دارد که از این استاندارد استفاده کرده باشند و برای عموم منتشر شده‌است[۲]. این دادگان جهت ساخت مدل‌های هوش مصنوعی مبتنی بر یادگیری عمیق و برای استفاده در سامانه‌های پرسش و پاسخ زبان فارسی‌است. نتایج این پژوهش نشان می‌دهد دادگان پرسش‌وپاسخ زبان فارسی ایجادشده می‌تواند پاسخ به سؤالات مطرح‌شده به زبان طبیعی فارسی را با معیار تطابق دقیق[۳] ۷۸ درصد و معیار F1 ۸۷ درصد برساند که هنوز نیازمند ارتقا است.

واژگان: دادگان پرسش‌وپاسخ زبان فارسی،سیستم‌های پرسش‌وپاسخ[۴]، درک مطلب[۵]، یادگیری عمیق، پردازش زبان طبیعی

Farsi Question and Answer Dataset (FarsiQuAD)

Javad ForutanRad*[1], Maryam Hourali [2], Mohammad Ali kevanRad[3]
[1] M.Sc in AI, Malek-Ashtar university of technology,
[2,3] Assistant professor in AI, electronic & computer Dep, Malek-Ashtar university of technology,

Abstract

A fast and accurate response to questions posed in natural language is a fundamental objective in the advancement of question and answer systems. These systems involve computers comprehending textual content and questions, and subsequently, delivering precise answers to users. Despite significant advancements in this field, there remains room for improvement, particularly when dealing with languages other than English, such as Persian.

In this article, we present the Persian language question and answer dataset, known as FarsiQuAD. This dataset was meticulously crafted by human annotators, drawing from Persian Wikipedia articles. FarsiQuAD is made available in two versions: Version 1 comprises over 10,000 questions and answers, while Version 2 offers an extensive collection of over 145,000 rows. This dataset is designed to seamlessly integrate with the English version of SQuAD and other databases in various languages adhering to this standard, and it is open to the public. These data serve as valuable resources for the development of

[1] Exact match
[2] https://github.com/Forutanrad/FarsiQuAD
[3] Exact match
[4] Question Answering
[5] Reading comprehension

* Corresponding author
*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۴۰۱/۶/۱۰ • تاریخ پذیرش: ۱۴۰۲/۹/۲۰ • تاریخ انتشار: ۱۴۰۲/۱۲/۲۹ • نوع مطالعه: بنیادی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

artificial intelligence models based on deep learning and for the enhancement of Persian language question and answer systems.
The research findings reveal that the FarsiQuAD dataset is capable of providing answers to questions posed in the natural Persian language with an exact matching accuracy of 78% and an F1 score of 87%. However, there is still room for improvement in achieving even higher accuracy levels.
This project arises from the critical need for non-English languages to have access to more data for training deep learning models, especially in the domain of factoid questions. Hence, the primary objective of this article is to introduce the newly created dataset. Prior to this effort, well-known datasets like SQuAD predominantly focused on English, and similar datasets has been developed in other languages, including French, German, Korean, and Japanese. Nevertheless, the dearth of question datasets in the Persian language was evident. The quality and diversity of questions are pivotal aspects, and as this dataset continues to grow, it will contribute to the broader landscape of research in this domain, allowing for valuable cross-linguistic comparisons and integration with research conducted in other languages.

Keywords: Question And Answer Dataset, Question And Answer systems, Reading omprehension, Deep Learning, Natural Language Processing, Factoid Questions

۱- مقدمه

پرسش‌وپاسخ یکی از مهم‌ترین زمینه‌های پژوهشی در بازیابی اطلاعات است و ترکیبی از دامنه‌های مختلف پژوهشی همچون پردازش زبان طبیعی، هوش مصنوعی، بازیابی اطلاعات و استخراج اطلاعات است که به‌طور خودکار به پرسش‌های زبان طبیعی با استفاده از منابعی که در پایگاه دانش ذخیره شده است، پاسخ می‌دهد. پاسخ‌گویی به سؤالات یکی از حوزه‌های بسیار پرکاربرد نیز هست که در بازیابی اطلاعات، سامانه‌های تعاملی، ربات‌ها، سامانه‌های کنترلی، نرم‌افزارهای گفتگو و هوش مکالمه‌ای کاربرد گسترده دارد. در این سامانه‌ها بسته به نوع سؤال، اطلاعات از پایگاه دانش استخراج و در اختیار قرار می‌گیرد و این موضوع باعث کاهش چشم‌گیر زمان دریافت پاسخ سؤالات و خودکارشدن فرایندهای مرتبط می‌شود.

با پیشرفت چشم‌گیر یادگیری عمیق در سالیان اخیر و افزایش دقت مدل‌های مبتنی بر این روش، امروزه بیشتر کارهای پرچالش و سخت پردازش زبان طبیعی از این رویکرد استفاده می‌کنند و شرکت‌های بزرگ دنیا مثل گوگل و فیس‌بوک نسبت به جایگزینی روش‌های سنتی با یادگیری عمیق به‌دلیل نتایج فوق‌العاده آن اقدام کرده‌اند. با اینکه کارهای زیادی به‌منظور توسعه مدل‌های پرسش‌وپاسخ برای سایر زبان‌های روز دنیا توسعه‌یافته‌اند، اما کمبود جدی برای زبان فارسی احساس می‌شود تا مشابه مدل‌های زبان انگلیسی توسعه یابند.

استفاده از یادگیری عمیق در مسائلی مانند زبان طبیعی که پیچیدگی‌های زیادی دارد، جزو پرکاربردترین و پیشرو این حوزه است. مهم‌ترین برتری شبکه‌های عصبی یادگیری عمیق این است که مهندسی ویژگی جداگانه را از بین می‌برند؛ لذا نیاز به برچسب‌گذاری، استخراج ویژگی و نیاز به افراد خبره زبان‌شناس را که جزوی از محدودیت‌های پروژه‌ها است، مرتفع می‌سازند و این کار توسط خود شبکه انجام می‌شود. دومین دلیل، تا حدود زیاد، مرتفع‌کردن محدودیت‌های معنایی است؛ باتوجه‌به اینکه مدل‌ها به‌طور معمول با چندین زبان مختلف آموزش می‌بینند و مفهوم کلمات و جملات در مدل قرار می‌گیرد؛ درنتیجه نتایج قابل قبولی را ارائه می‌کنند و برتری سوم سرعت مدل‌ها است که نسبت به مدل‌های و روش‌های سنتی سرعت اجرای به‌مراتب بالاتری دارند؛ اما مشکل اصلی در مدل‌های یادگیری عمیق، نیازمندی مدل‌ها به داده‌های اولیه خیلی زیاد است.

۲- تاریخچه

۲-۱- کارهای انجام‌شده در زبان انگلیسی

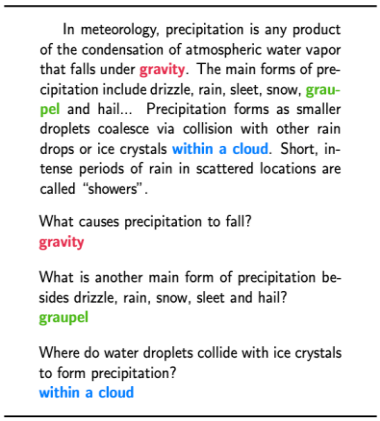
برای زبان انگلیسی کامل‌ترین دادگان منتشرشده دادگان SQuAD به زبان انگلیسی در دو نسخه با شماره ۱.۰ [۱] و ۲.۰ [۲] است که توسط دانشگاه استنفورد منتشر شده‌است.

مجموعه‌داده پاسخ به سؤالات استنفورد (SQuAD)، مجموعه‌داده‌های درک مطب بیش از صدهزار سؤال است که توسط اعضای گروه از مقالات ویکی‌پدیا تهیه شد و در آن پاسخ به هر سؤال، بخشی از متن مربوطه است[۱].

این مجموعه برای کمک به چالش‌های پرسش‌وپاسخ در NLP است. این دادگان بر وظیفه پاسخ به سؤالات زبان طبیعی تمرکز دارد. در شکل (۱) نمونه‌ای از جفت پرسش‌وپاسخ‌هایی که توسط سازندگان ارائه شده گذاشته شده‌است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]


In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

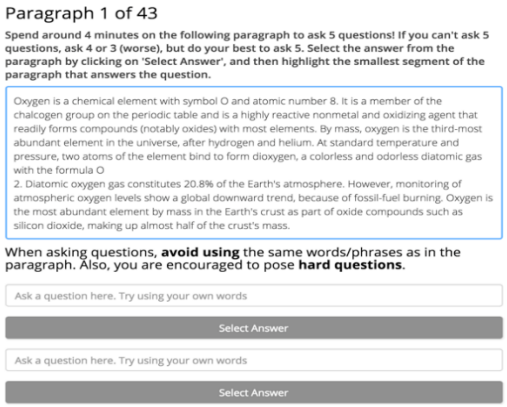
Where do water droplets collide with ice crystals to form precipitation?
within a cloud

(شکل – ۱) – **نمونه جفت پرسش‌وپاسخ ارائه شده در ویکی‌پدیا**[۱]

(Figure -1) - A sample question-and-answer pair provided on Wikipedia

در نسخه ۱ تعداد ۵۳۶ مقاله را از بین ۱۰۰۰۰ مقاله ویکی‌پدیا نمونه‌برداری کرده‌اند که در مجموع ۲۳۲۱۵ پاراگراف جداگانه استخراج شده‌است. در این پاراگراف‌ها مواردی که کمتر از پانصد نویسه و بیشتر از ۱۵۰۰ نویسه داشته‌اند، حذف شده‌اند. در این دادگان هشتاددرصد مقالات را به مجموعه آموزشی، ده درصد را به مجموعه آموزشی و ده درصد را به مجموعه آزمون داده‌ها اختصاص داده‌اند.

در این مجموعه از استخراج‌کنندگان سؤال خواسته شده‌است که از هر مقاله پنج سؤال و یا حداقل سه سؤال استخراج کند، به‌طوری‌که حتماً جواب سؤال در متن مقاله موجود بوده و از داخل پاراگراف مقاله متن جواب سؤال را برجسته کرده‌اند. برای این منظور نرم‌افزاری در اختیار استخراج‌کنندگان سؤالات گذاشته شده‌است که می‌توانستند سؤالات را بدون رونوشت از متن، وارد و سپس پاسخ را از روی متن انتخاب کنند.


Paragraph 1 of 43

Spend around 4 minutes on the following paragraph to ask 5 questions! If you can't ask 5 questions, ask 4 or 3 (worse), but do your best to ask 5. Select the answer from the paragraph by clicking on 'Select Answer', and then highlight the smallest segment of the paragraph that answers the question.

Oxygen is a chemical element with symbol O and atomic number 8. It is a member of the chalcogen group on the periodic table and is a highly reactive nonmetal and oxidizing agent that readily forms compounds (notably oxides) with most elements. By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O
2. Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere. However, monitoring of atmospheric oxygen levels show a global downward trend, because of fossil-fuel burning. Oxygen is the most abundant element by mass in the Earth's crust as part of oxide compounds such as silicon dioxide, making up almost half of the crust's mass.

When asking questions, **avoid using** the same words/phrases as in the paragraph. Also, you are encouraged to pose **hard questions**.

Ask a question here. Try using your own words

Select Answer

Ask a question here. Try using your own words

Select Answer

(شکل – ۲). نمونه روش مورداستفاده برای جمع‌آوری سؤالات دادگان SQuAD [۱]

(Figure - 2) - The method used for collecting questions in the SQuAD dataset[۱]

در سایر سؤالات مبتنی بر سند که به مجموعه‌داده‌ها پاسخ می‌دهند و بر استخراج پاسخ تمرکز می‌کنند، پاسخ به یک سؤال دادهشده در چندین سند وجود دارد؛ بااین‌حال، در SQuAD، این مدل فقط به یک گذرگاه واحد دسترسی دارد و کار بسیار دشواری را ارائه می‌دهد.

یک نوع مجموعه‌داده محبوب، مجموعه‌داده cloze است که از یک مدل می‌خواهد کلمه‌ای را که در یک قسمت وجود ندارد، پیش‌بینی کند. این مجموعه‌داده‌ها بزرگ هستند و وظیفه‌ای تا حدودی مشابه SQuAD را ارائه می‌دهند. پیشرفت کلیدی که SQuAD در این زمینه انجام می‌دهد این است که پاسخ‌های آن پیچیده‌تر هستند؛ و بنابراین نیاز به استدلال بیشتری دارند؛ بنابراین SQuAD را برای ارزیابی درک و قابلیت مدل بهتر می‌کند.

در نسخه ۲.۰ این دادگان[۲] تعداد زیادی سؤال از نوع نداشتن پاسخ افزوده شده‌است. در این نوع سؤال‌ها سامانه می‌تواند متوجه شود که پاسخ را نمی‌داند.

نسخه ۲.۰ دادگان بیش از ۱۵۰۰۰۰ جفت پرسش‌وپاسخ دارد.[۲].

۲-۲- بررسی دادگان پرسش‌وپاسخ به سایر زبان‌ها غیر از انگلیسی

باتوجه‌به اینکه شبکه‌های آموزش دیده‌شده با مجموعه سؤالات به زبان مقصد نتایج بهتری نسبت به مدل‌های چندزبانه داشته‌اند در زبان‌های مختلف این مجموعه سؤالات بر مبنای سؤالات دادگان SQuAD تهیه شده‌است.

برای زبان فرانسوی دادگان FQuAD [3] تهیه شده که دارای بیش از ۶۰،۰۰۰ جفت پرسش‌وپاسخ است. برای زبان آلمانی دادگان GermanQuAD [4] ارائه شده که شامل 13,722 جفت پرسش‌وپاسخ‌است. برای زبان کره‌ای دو کار [5,6] انجام شده‌است و برای زبان ژاپنی دادگان JaQuAD [7] که مشتمل بر 39,696 هزار جفت پرسش‌وپاسخ‌است. در زبان فارسی دادگان PersianQA [8] منتشر شده‌است که ده‌هزار جفت پرسش‌وپاسخ دارد که به نسبت سایر زبان‌ها تعداد کمتری است.

۳- مدل پیشنهادی

یکی از مهم‌ترین قسمت‌های شبکه یادگیری عمیق نیازمند بودن به داده‌های زیاد اولیه است که گاهی اوقات

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

به‌عنوان محدودیت این شبکه‌ها نیز به دلیلی سختی تهیه این داده‌ها نیز محسوب می‌شود. سختی ساخت دادگان باعث شده‌است که افراد و شرکت‌های محدودی در این زمینه فعالیت کرده و یا بعضاً نتایج و دادگان خودشان را در اختیار سایرین قرار ندهند.

در این کار تلاش شده‌است با الگوبرداری از نمونه زبان انگلیسی دادگان دانشگاه استنفورد (SQuAD) که برای پاسخ‌گویی به پرسش‌وپاسخ از متون به‌صورت حقیقت‌نما و به زبان انگلیسی تهیه شده، برای زبان فارسی نیز توسعه یابد.

۳-۱- بررسی ساختارها و داده‌های خام اولیه

به‌منظور تهیه مجموعه نیاز بود مقالات پایه اولیه جهت استخراج سؤال‌ها و پاسخ‌ها شناسایی و انتخاب شوند.

مقالات ویکی پدیا [۱۷] این دادگان شامل تمام مقالات خام ویکی‌پدیای فارسی تا تاریخ ۱۲ مرداد ۱۳۹۹ است این دادگان به‌صورت ۹ عدد فایل‌های متنی حجیم ارائه شده‌است. این دادگان شامل ۷۳۹۸۷۰ مقاله، ۴۰۰۴۷۶۵ جمله و ۹۴۰۰۲۰۹۴ کلمه‌است.

نظر به بررسی‌های صورت‌گرفته در مجموعه‌های مختلف دادگان‌های بررسی‌شده، به دلیل اینکه دادگان ویکی‌پدیا [۱۷] دارای مقاله‌هایی در حوزه‌های مختلف، به‌روزتر از سایر مجموعه‌داده‌ها بود، لذا برای استفاده در سامانهٔ پرسش و پاسخ مناسب‌تر تشخص داده شد.

گفتنی است دادگان SQuAD نیز از مقالات ویکی‌پدیا برای استخراج سؤالات استفاده کرده است که این موضوع باعث اعتبار بیشتر به این مجموعه و انتخاب آن شد.

در این کار سعی شده‌است که از مقالات عمومی ویکی‌پدیا استفاده شود؛ لذا از میان ۶۳۲۶۴ به‌صورت تصادفی و به انتخاب افراد خبرهٔ استخراج‌کنندهٔ سؤالات، تعداد ۲۷۰۵ مقاله مورد استفاده قرار گرفته که در سی گروه مختلف اعلام‌شده در مقاله بوده که تنوع زیادی را شامل شده‌است.

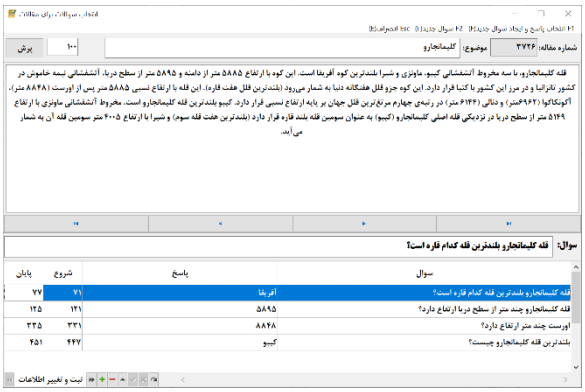
۳-۱-۱-آماده‌سازی داده‌ها و مرتب‌کردن

باتوجه‌به اینکه داده‌های دادگان ویکی‌پدیا به‌صورت فایل‌های متنی خام سنگین ارائه شده بود، لذا برای آماده‌سازی جهت استفاده با استفاده از کد زبان برنامه‌نویسی پایتون هر ۹ فایل متنی دادگان پردازش و مقالاتی که تعداد نویسه‌های آن بین ۵۰۰ الی ۱۱۰۰ نویسه بودند، استخراج شدند و سپس با استفاده از کتابخانه هضم عملیات، پیش‌پردازش مربوط به تمیز و مرتب‌کردن و نرمال‌سازی داده صورت گرفت. در این مرحله حرف "ه" که اشکالاتی داشت، جایگزین شد.

۳-۱-۲- ساخت برنامه جهت تسهیل استخراج سؤالات

به‌منظور جمع‌آوری و استخراج سؤالات، و همچنین هماهنگی، نیاز به یک واسط کاربری قوی، سریع و کاربرپسند بود تا بتوان به‌سادگی سؤالات موردنیاز را استخراج کرد؛ لذا برای این منظور نرم‌افزاری به شرح زیر ایجاد شد. نام این نرم‌افزار QAProject (انتخاب سؤالات برای مقالات) نام‌گذاری شد.

این برنامه با اتصال به بانک اطلاعاتی مقالات را واکشی و اطلاعات شماره مقاله، موضوع و متن را نمایش داده و کاربر می‌تواند در قسمت سؤال، متن سؤالی را که جواب آن در متن است، وارد کرده، سپس قسمت مرتبط به پاسخ را از متن با انتخاب قسمت درست پاسخ مشخص کرده و با فشردن کلید F1 قسمت انتخاب‌شده از متن به‌عنوان پاسخ انتخاب شده و به همراه نویسه شروع و پایان در بانک اطلاعاتی ذخیره شده و یک ردیف جدید برای سؤال بعدی ایجاد می‌شود.



(شکل - ۳). نمایی از برنامه ساخته‌شده برای استخراج سؤالات

(Figure-3). Representation of the program created for question extraction

همان‌طور که در شکل (۵) نشان داده‌شده‌است، مقالات در بالای صفحه و قسمت سؤالات و پاسخ‌ها در پایین صفحه قرار گرفته‌است. در قسمت سؤالات امکان ویرایش، حذف، درج جدید و به‌روزرسانی و در قسمت سمت چپ بالای نرم‌افزار امکان پرش به تعداد زیادی مقاله افزوده شده‌است و در مثال بالا با هر بار زدن روی دکمه پرش،

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.4.107]

نرم‌افزار صد مقاله جلوتر می‌رود. در قسمت وسط صفحه امکان رفتن به مقاله بعدی، قبلی، آخرین و اولین نیز گذاشته شده‌است.

این نرم‌افزار به همراه تقسیم مساوی از تعداد کل مقالات با کنترل اینکه از قبل در اختیار شخص دیگری نبوده و تضمین عدم دریافت مقاله تکراری، در اختیار کارشناسان استخراج‌کننده سؤال که در این پروژه هفت نفر بودند، قرار داده است، شخص بعد از مطالعه کامل متن مقاله، در قسمت سؤال، متن سؤالی را که در مقاله جواب آن موجود است، وارد کرده و سپس برای جواب آن نیز باید قسمتی از متن مقاله در اختیار را با موش‌واره انتخاب و سپس تأیید کند تا به‌عنوان جواب سؤال درج شود و امکان درج متن تایپی برای پاسخ مقدور نبوده‌است. گفتنی است هم‌زمان با انتخاب جواب، نویسه شروع و پایان جواب از مقاله نیز استخراج و ذخیره شده‌است. ضریب کاپا ۰.۹۸۳ درصد بوده‌است.

۳-۱-۳- تهیه سؤالات

به‌منظور در ادامه نسبت به تهیه پرسش‌وپاسخ به زبان فارسی با استفاده از نرم‌افزار ایجادشده اقدام شد.

به‌منظور استخراج سؤالات، تیمی هفت نفره از افراد خبره با مدرک تحصیلی کارشناسی‌ارشد فعالیت داشته‌اند.

- **نمونه سؤالات استخراج شده**

در ادامه نمونه‌ای از متن مقاله و سؤال‌های پرسیده‌شده به همراه پاسخ استخراج‌شده به نمایش گذاشته شده‌است.

متن مقاله: " یغوارد (به ارمنی: ԵՂՎԱՐԴ) یک شهر در غرب ارمنستان است که در استان کوتایک واقع شده‌است. در کیلومتری شمال هرازدان، مرکز استان کوتایک واقع شده‌است. کلمه یغوارد از دو کلمه ارمنی یغی () به معنی (عطر یا بو) و وارد () به معنی (گل رز) مشتق شده‌است. به عقیده آرام غانالانیان پژوهش‌گر و ارمنی‌شناس نام‌گذاری یغوارد به‌دلیل وجود یک جنگل وسیع در این منطقه است که از انواع گل رز و سایر گل‌های با عطر زیاد پوشیده شده‌است یغوارد از کهن‌ترین اقامتگاه‌های بشری در منطقه است و در آن آثار معماری زیادی یافت می‌شود. یغوارد **۷ کیلومتر مربع** مساحت دارد. یغوارد در جنوب کوه آرایی در ارتفاع ۱۳۳۳ متری از سطح دریا واقع شده‌است. این شهر در **۱۵ کیلومتری** ایروان پایتخت جمهوری ارمنستان قرار گرفته است کلیسای یغوارد بنا شده در ۱۳۰۱ تنها بنای حفظ‌شده تاریخی در یغوارداست. اف سی یغوارد یک باشگاه فوتبال بود و نماینده این شهرک در بین سال‌های ۱۹۸۶ تا ۱۹۹۶ بود و به دلایل مشکلات مالی منحل شد.

سؤال ۱: یغوارد در کدام کشور قرار دارد؟ ارمنستان

سؤال۲: شهر یغوارد در کدام قسمت از ارمنستان واقع است؟ غرب

سؤال ۳: شهر یغوارد در کدام استان قرار دارد؟ پاسخ: کوتایک

سؤال ۴: یغوارد چقدر مساحت دارد؟ پاسخ: **۷ کیلومتر مربع**

سؤال ۵: شهر یغوارد در چه فاصله ای از ایروان واقع است؟ پاسخ: **۱۵ کیلومتری**"

همان‌طور که در سؤالات استخراج شده از این مقاله مشخص است در قسمت مساحت چون عنوان نشده بود که به چه واحدی نمایش داده شود پس کلمات بعد از عدد هم جزوی از سؤال‌است. مثلاً در سؤال اگر گفته شده بود نتیجه سؤال مساحت چند کیلومترمربع است در جواب فقط عدد نمایش داده خواهد شد.

- **خلاصه آماری**

تعداد کل مقالات ۶۳۲۶۴ مقاله بوده‌است که ازاین‌بین، تعداد ۲۷۰۵ مقاله به‌عنوان مقاله منتخب از حوزه‌های محتلف موجود انتخاب و سؤالات از این مقالات استخراج شده‌اند.

تعداد کل سؤالات استخراج‌شده بیش از ده‌هزار جفت پرسش‌وپاسخ است که با درک مطلب و به‌صورت مفهومی از متن استخراج شده‌است. در این مجموعه سؤالات دقت شده است که جواب سؤال باید در متن قابل استخراج باشد و برای سؤالاتی که سامانه باید بداند که جواب آن را نمی‌داند نیز ۱۵۳ سؤال در نسخه ۱ ثبت شده‌است؛ لذا تعداد سؤالات حاوی پاسخ ۹۸۴۷ جفت هستند.

در این دادگان کمترین تعداد سؤال استخراج‌شده از مقاله‌ها عدد ۱، بیشترین تعداد سؤال استخراج‌شده از مقاله ۱۲ و میانگین سؤال استخراج‌شده از هر مقاله ۳.۷ است.

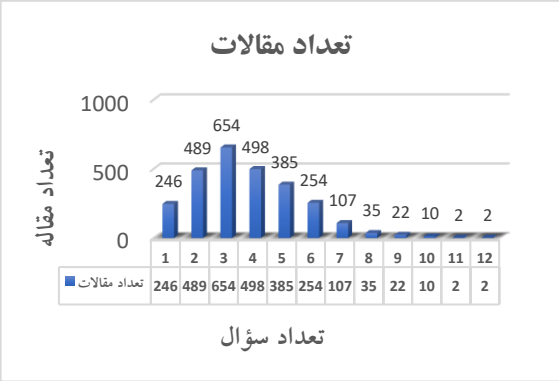
در شکل (۶) فراوانی مقالات بر اساس تعداد سؤال‌هایی که از هر مقاله استخراج شده، نمایش داده‌شده‌است. همان‌طور که در شکل مشخص است بیشترین استخراج سؤال مربوط به استخراج ۳ سؤال از

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

مقاله‌ها بوده که ۶۵۴ مقاله را شامل می‌شود و رتبه دوم و سوم به‌ترتیب در اختیار ۴ و ۲ سؤال از مقاله‌ها قرار دارد.

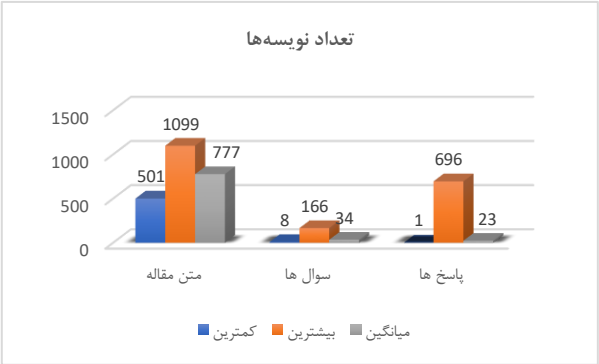
برای استخراج سؤالات این دادگان به‌صورت متوسط می‌توان ۴۰ تا ۵۰ سؤال را در یک ساعت استخراج کرد که با توجه به تعداد ده‌هزار سؤال استخراج‌شده برابر با ۲۵۰ ساعت زمان صرف استخراج سؤالات از متون مقاله‌ها شده‌است.



(شکل – ۴). فراوانی مقالات در تعداد سؤال استخراج شده

(Figure-4) - Frequency of articles in the number of extracted questions

در شکل (۷) تعداد نویسه‌های مقالات، سؤالات استخراج شده و پاسخ‌ها به‌صورت گزارش کامل ارائه شده‌است. در این نمودار مشخص می‌شود که مقالات به‌صورت متوسط ۷۷۷ نویسه داشته‌اند، پاسخ‌ها به‌صورت میانگین ۲۳ نویسه و سؤال‌های مطرح‌شده نیز به‌صورت میانگین دارای ۳۴ نویسه بوده‌اند.



(شکل –۵). نمودار تعداد نویسه‌های مقالات، سؤالات و پاسخ‌ها

(Figure-5). Chart of the number of characters of articles, questions and answers

- **موضوع مقالات استفاده شده در تهیه دادگان**

موضوعات استفاده شده دارای طیف گوناگونی از موضوعات است که در ذیل تعدادی از این موارد مطرح شده‌است:

اماکن و موزه‌ها، حروف الفبا، اشخاص و شخصیت‌ها، ادبیات، کتاب و رمان، موسیقی، آلات موسیقی و نوازنده‌ها، وسایل جنگی، ریاضیات، جغرافیای کشورها، استان‌ها، شهرستان‌ها، شهرها و روستاها، تاریخی، عناصر شیمیایی، برق، الکترونیک و رایانه، اقتصاد و یکای پولی، فیلم و سینما، قرآن، احزاب، حیوانات، گل، درختان و گیاهان، ورزشی و ورزش‌کاران، بزرگراه‌ها، آسمان، ستاره، کهکشان و صورت‌های فلکی‌ها

- **سؤالاتی که باید بداند که جواب آن را نمی‌داند**

خیلی مهم است که سامانه‌ها بدانند، جواب بعضی از سؤالات را با این‌که تمام یا بخش زیادی از کلمات در متن وجود دارند، نمی‌دانند. به‌عنوان‌مثال در متن آمده است "جمعیت شهر بیرجند ۲۱۰۰۰۰ هزار نفراست." در این متن می‌توان به سؤال "جمعیت بیرجند چند نفر است؟" پاسخ دهد؛ اما به سؤال "شهر بیرجند چند نفر مرد جمعیت دارد؟" نمی‌تواند پاسخ دهد چون همچنین چیزی در متن عنوان نشده‌است و فقط کلمه "مرد" در متن سؤال باعث می‌شود که دیگر امکان پاسخ‌گویی به سؤال وجود نداشته باشد.

این حوزه شامل بخش زیادی از سؤالات مختلف می‌شود که انسان این موضوع را می‌داند البته به‌سختی قابل تشخیص است و یکی از چالش‌های شبکه‌های هوش مصنوعی در پاسخ‌گویی به سؤالات هستند.

- **مواردی که در تهیه سؤالات رعایت شده‌است:**
  - در انتخاب پاسخ تلاش شده حداقل‌ترین پاسخ ممکن انتخاب شود.
  - در تهیه پاسخ‌هایی که با پیشوندهایی مثل خانم، آقای، حاجی، مهندس، ژنرال و... بوده‌اند به‌عنوان جزوی از جواب انتخاب شده‌است.
  - پیشوندهایی مثل سال، شهر و ... درصورتی‌که در متن سؤال قرار داشتند در جواب در نظر گرفته نشده‌اند.
  - پسوندهایی مانند متر، میلادی، کیلومتر، میلیون، قمری، خورشیدی، میلی‌متر، و... نیز جزوی از جواب در نظر گرفته شده البته درصورتی‌که در خود متن سؤال این کلمات وجود نداشته‌اند.
  - کلماتی که بیان‌گر تقریبی هستند، مثل حدوداً، تقریباً، بیش از نزدیک به و... اگر در صورت

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

سؤال عنوان نشده‌اند در جواب به‌عنوان بخشی از پاسخ در نظر گرفته شده‌اند.

- انعکاس کامل متنی که در نقل‌قول استفاده شده‌است.
- در استخراج تلاش شده‌است از نوع سؤالات مختلف شامل مالکیتی، زمان، چرایی، کجایی، چیستی، چگونگی، چطوری، چرا، کدام و ... استفاده شود.
- در تهیه سؤالات از سؤالاتی با چندین ارتباط استفاده شده‌است؛ برای مثال در سؤال گفته شده‌است "شغل پدر نویسنده مقاله چیست؟"، درحالی‌که در متن مقاله اسم شخص آمده و پدر شخص عنوان شده و در ادامه به شغل پدر او پرداخته شده‌است.
- سؤالاتی شرطی نیز استخراج شده‌است مثل سؤال "شرکت ایران‌خودرو دولتی است یا خصوصی"؟
- در تهیه سؤالات از کلمات معادل و یا مفهومی که در متن دقیقاً استفاده نشده، به تکرار استفاده شده‌است.
- در تهیه سؤالات هم در متن سؤال و هم در جواب سؤال از کلماتی که به‌غیراز زبان فارسی بوده‌اند نیز استفاده شده‌است. برای مثال سؤال " سنبل به کدام تیره تعلق دارد؟ " جواب: " Hyacinthaceae".
- در سؤال‌وجواب آن از معادل نوشتار اعداد فارسی و برعکس آن استفاده شده‌است؛ مثل (سه، دو، یک و ...)
- در طراحی سؤال‌ها دقت شده که هر سؤال فقط از یک مقاله قابل استخراج و در مقالات دیگر احتمال داشتن پاسخ خیلی پایین باشد.

۳-۱-۴-آماده‌سازی دادگان

در این مرحله داده‌های ایجادشده به دو قسمت آموزش و آزمون تقسیم شدند، در این مرحله برای نسخه یک تعداد نه‌هزار سؤال برای داده‌های آموزش و هزار سؤال برای داده‌های آزمون در نظر گرفته شدند و فایل‌های JSON مطابق استاندارد SQuAD ایجاد شدند. این فایل‌ها قابلیت ادغام با نسخه اصلی SQuAD را دارند.


```
{
    "data": [
        {
            "title": "کمال خجندی",
            "paragraphs": [
                {
                    "qas": [
                        {
                            "answers": [
                                {
                                    "answer_start": ۵۵,
                                    "answer_end": ۹۷,
                                    "text": "از عارفان و شاعران پارسی گوی قرن هشتم هجری"
                                }
                            ],
                            "question": "کمالی خجندی که بود؟",
                            "is_impossible": false,
                            "id": "b۵۹da۵۹۸-adf۵-۴e۲c-aa۴۷-۳۲e۲۶۳d۰f۴۵d"
                        },
                        {
                            "answers": [
                                {
                                    "answer_start": ۱۱۲,
                                    "answer_end": ۱۲۳,
                                    "text": "خجند فرارود"
                                }
                            ],
                            "question": "کمال خجندی در کجا به دنیا آمد؟",
                            "is_impossible": false,
                            "id": "d۴c۲f۲ab-۸c۱۱-۴۹۲d-b۵da-۵c۰۵۲e۲b۲۷b۶"
                        }
                    ],
                    "context": " آرمگاه او در تبریز است. این بیت بر لوح آرامگاه او ثبت شده است: دیوان کمال خجندی در وبگاه گنجور "
                }
            ]
        },[
```


(شکل-۶). نمونه ساختار فایل دادگان FarsiQuAD

(Figure-6). Sample structure of the FarsiQuAD dataset file

۳-۱-۵- نسخه شماره دو برای دادگان

با بررسی‌هایی که در مورد نسخه یک دادگان انجام شده نتایج حاصل برای یک مقاله قابل‌قبول بود، اما وقتی جستجو در تعداد زیادی از مقالات انجام می‌شد، باعث شده بود نتایج حاصل به‌دلیل اینکه قبلاً این متون برای سؤال دیده نشده بودند، به‌عنوان پاسخ نمایش داده شوند؛ لذا برای غلبه بر این مشکل دادگان نسخه شماره ۲ (FarsiQuAD_V2) تهیه شد. در این دادگان به‌صورت تصادفی پنجاه سؤال را برای هر مقاله که مربوط به این مقاله نیستند انتخاب، و به‌عنوان سؤالاتی که باید مدل‌ها بدانند جواب آن را نمی‌دانند (جواب غیرممکن) به دادگان افزوده شد. با انجام این کار تعداد ردیف‌های دادگان به بیش از ۱۴۵۰۰۰ سؤال رسید.

۳-۱-۵- مقایسه دو دادگان پرسش‌وپاسخ زبان فارسی

در این قسمت به شباهت‌ها و تفاوت‌های دو دادگان پرسش‌وپاسخ زبان فارسی FarsiQuAD و PersianQA پرداخته می‌شود.

شباهت‌ها در دو دادگان به شرح زیر است:

- بر اساس درک مطلب تهیه شده‌اند.
- بر اساس ساختار دادگان معروف SQuAD تهیه شده‌اند و قابلیت ادغام در نسخه اصلی را دارند.
- به زبان فارسی هستند.
- دارای سؤال‌های فاقد جواب هستند.
- منبع مقالات جهت استخراج پاسخ‌ها ویکی‌پدیای فارسی‌است.
- تعداد جفت‌سؤال و پاسخ‌ها تقریباً یک اندازه هستند.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

تفاوت‌های عمده دو دادگان به شرح زیر است:

- دادگان PersianQA محاوره‌ای است و FarsiQuAD نوشتار رسمی دارد.
- در PersianQA جواب‌ها از چندین مقاله قابل استخراج است در مقابل FarsiQuAD سؤال‌ها از یک مقاله استخراج می‌شود.
- تعداد سؤالات استخراج‌شده از هر مقاله با هم متفاوت است.
- میانگین طول نویسههای مقاله‌های FarsiQuAD تقریباً سه برابر مقاله‌های PersianQA است.
- میانگین طول نویسه‌های سؤال‌های FarsiQuAD حدوداً چهار برابراست.
- میانگین طول نویسه‌های پاسخ‌های FarsiQuAD تقریباً بیش از دوبرابر است.

(جدول – ۱). مقایسه FarsiQuAD با PersianQA

(Table –1) Comparison of FarsiQuAD with PersianQA

مقایسه نویسهها	Farsi QuAD V1	Farsi QuAD V2	Persian QA
تعداد جفت پرسش‌وپاسخ	۱۰۰۲۱	۱۴۵۶۲۲	۹۹۳۸
بدون پاسخ	152	۱۳۵۷۵۳	2980
میانگین طول سؤالات	۳۴	۳۴	۸.۳۹
میانگین طول مقاله‌ها	۷۷۷	۷۷۷	۲۲۴.۵۸
میانگین طول پاسخ	۲۳	۲۳	۹.۶۱

۳-۱-۷-آموزش مدل یادگیری عمیق

در این بخش مدل‌های مختلف شبکه‌های یادگیری عمیق را با استفاده از داده‌هایی که در گام قبل آماده کرده بودیم آموزش داده شد که نحوه آموزش و ارائه به شرح زیر است:

به‌منظور آموزش از مدل‌های از پیش آموزش‌دیده شده و روش یادگیری انتقالی استفاده شده تا بتوان مدل‌های که از قبل آموزش‌دیده شده‌اند برای انجام‌وظیفه پرسش‌وپاسخ به زبان فارسی تنظیم د.

برای آموزش مدل‌ها از کتابخانه پایتورچ[1] ، کتابخانه مبدل[2] huggingface، CUDA و اسکریپت ارائه شده برای آموزش مدل های SQuAD استاندارد[3] استفاده شد تا آموزش و نتایج آن استاندارد باشد.

برای آموزش از سخت‌افزار GPU RTX2070 و شبکه‌های مختلف با مدت بیش از ۸۵ ساعت آموزش دیدند.

[1] https://pytorch.org
[2] https://github.com/huggingface/transformers
[3]https://github.com/huggingface/transformers/blob/main/examples/legacy/question-answering/run_squad.py

به‌منظور آموزش مدل‌ها دستور ذیل و با پارامترهای نمونه اجرا شده‌است:

```
python ./src/run_squad.py \
    --model_type bert \
    --model_name_or_path "${model}" \
    --do_train \
    --do_eval \
    --train_file $DATA_DIR/FarsiQA_V1_Train.json \
    --predict_file $DATA_DIR/FarsiQA_V1_Eval.json \
    --learning_rate "${learning_rate[@]}" \
    --num_train_epochs "${num_train_epoch[@]}" \
    --max_seq_length 384 \
    --doc_stride 128 \
    --output_dir
"Question&Answer_JFmodel/${model}_learning_rate=${learning_rate}_num_t
rain_epoch=${num_train_epoch}" \
    --per_gpu_eval_batch_size=256  \
    --per_gpu_train_batch_size=4   \
    --save_steps 5000
```

(شکل – ۷). نمونه اسکریپت اجرا شده برای آموزش مدل

(Figure-7). Sample script executed for model training

۳-۱-۸-آموزش مدل‌ها با نسخه ۱ دادگان (FarsiQuAD)

برای شروع آموزش‌ها، مدل‌ها با هایپر پارامترهای ذیل تنظیم و برای داده‌های دادگان ساخته شده آموزش دیدند.

(جدول-۲) تنظیم هایپرپارامترها برای آموزش مدل‌ها

(Table - ۲). Hyperparameter Settings for Model Training

max_seq_length	۳۸۴
doc_stride	۱۲۸
Batch_Size	۴
نرخ یادگیری[4]	3e-5 و 5e-5
تعداد مراحل آموزش[5]	۳ و ۷

در این مرحله مدل‌های زیر با هایپرپارامترهای عنوان‌شده آموزش داده‌شده و برای وظیفه پرسش‌وپاسخ به زبان فارسی تنظیم شدند:

- bert-base-multilingual-uncased
- distilbert-base-uncased
- bert-base-multilingual-cased-finetuned-dutch-squad2[6]
- bert-base-parsbert-uncased
- bert-fa-base-uncased
- bert-fa-zwnj-base
- distilbert-fa-zwnj-base

۳-۱-۹-آموزش با دو مجموعه سؤالات زبان فارسی

با توجه اینکه مدل‌های ارائه شده بتوانند نمونه داده‌های بیشتری را ببینند تعدادی از مدل‌هایی که در مرحله قبل نتایج بهتری را به دست آورده بودند با استفاده از تجمیع فایل‌های آموزش و تست برای پیکره ایجاد شده FarsiQuAD و پیکره PersianQA ، مجدد آموزش داده شدند.

[4] Learning Rate
[5] epoch
[6] https://huggingface.co/henryk/bert-base-multilingual-cased-finetuned-dutch-squad2
این مدل قبلا توسط پیکره squad2 نیز آموزش داده شده‌است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.4.107]

آموزش مدل با تنظیم هایپر پارامترهای عنوان شده در جدول (۵) انجام شده‌است.

(جدول - ۳). تنظیم هایپرپارمترها برای آموزش دو مجموعه سؤالات زبان فارسی

(Table - ۳)..Hyperparameter Settings for Training Two Persian Language Question Sets

max_seq_length	۵۱۲
doc_stride	۲۵۶
Batch_Size	۴
نرخ یادگیری[1]	3e-5 و 5e-5
تعداد مراحل آموزش[2]	۳ و ۷

۳-۱-۱۰-آموزش با سه مجموعه دادگان پرسش‌وپاسخ

به‌منظور اینکه بررسی شود آیا آموزش مدل چندزبانه می‌تواند به بهبود عملکرد نتایج داده‌های تست زبان فارسی کمک کند یا خیر، از ترکیب سه مجموعه‌داده SQuAD V2.0[3]، مجموعه دادگان FarsiQuAD و دادگان PersianQA با هم ترکیب شدند تا مجدد شبکه آموزش ببینند. این ترکیب باعث شده شبکه به سنگین شود و مدت ۱۲ ساعت آموزش با GPU زمان برد تا آموزش انجام شود.

مدل پایه انتخاب شده برای آموزش به شرح ذیل‌است:

- bert-base-multilingual-uncased

نرخ یادگیری مدل 3e-5 و تعداد مراحل آموزش ۳ تنظیم شدند بقیه موارد تنظیمی مشابه مرحله قبل بوده‌است.

۳-۱-۱۱-آموزش با نسخه ۲ دادگان FarsiQuAD

به‌منظور افزایش دقت مدل‌ها برای مقابله با یافتن جواب‌های غیرقابل‌قبول، از تجمیع دو دادگان فارسی نسخه ۲ FarsiQuAD و دادگان PersianQA استفاده شد و با تنظیم هایپرپارامترهای نرخ یادگیری 3e-5 و تعداد مراحل آموزش ۳ مدل ذیل آموزش دید.

- bert-base-multilingual-cased-finetuned-dutch-squad2

این مدل کمک می‌کند وقتی جواب سؤال در متن وجود ندارد از آن متن برای یافتن پاسخ استفاده نشده و در پاسخ‌ها قرار نگیرند.

۳-۱-۱۲- ارزیابی مدل‌ها

به‌منظور ارزیابی نتایج ۱۰ درصد از سؤالات دادگان به‌صورت تصادفی به‌عنوان داده‌های تست در نظر گرفته شد که تعداد آن برای نسخه ۱ برابر ۱۰۰۰ جفت پرسش‌وپاسخ است و هایپرپارامترهای این مرحله متناسب با هایپرپارامترهای زمان آموزش تنظیم شدند.

اسکریپت ذیل برای تمام مدل‌های آموزش‌دیده شده اجرا شدند و خروجی بر اساس معیارهای ارزیابی ارائه شد که در قسمت ۵ به‌تفصیل نتایج ارائه شده‌است.

```
export model="My Model"
python ./src/run_squad.py \
  --model_type bert \
  --model_name_or_path "${model}" \
  --do_eval \
  --predict_file $DATA_DIR/eval.json \
  --max_seq_length 512 \
  --doc_stride 256  \
  --output_dir "${model}/eval" \
  --per_gpu_eval_batch_size=256
```

۴- ارزیابی و تجزیه‌وتحلیل نتایج

در این بخش به نتایج حاصل از کار انجام‌گرفته پرداخته شده و نتایج خروجی‌ها و آزمایش‌های انجام‌گرفته مورد ارزیابی و بررسی قرار خواهد گرفت.

۴-۱- معرفی معیارهای ارزیابی

دو معیار برای بسیاری از مجموعه‌های داده برای پاسخگویی به سؤالات استفاده می‌شود، از جمله این موارد تطابق دقیق (EM) و امتیاز F1است. این درصدها بر اساس پرسش‌های فردی + جفت پاسخ محاسبه می‌گردند. زمانی که چندین پاسخ صحیح برای یک سؤال امکان‌پذیر می‌شود، حداکثر درصد امتیاز نسبت به همه پاسخ‌های صحیح احتمالی محاسبه می‌شود. به‌طورکلی درصد امتیاز تطابق دقیق و F1 برای یک مدل با میانگین امتیازهای نمونه‌های فردی محاسبه می‌شود.

۴-۱-۱- تطابق دقیق

این معیار خیلی ساده‌است. برای هر جفت سؤال+پاسخ، اگر نویسههای پیش‌بینی مدل، دقیقاً با نویسههای (یکی از) پاسخ‌های واقعی مطابقت داشته باشد EM=1 و در غیر این صورت EM=0 می‌شود. این یک معیار دقیق همه یا هیچ است، یعنی با یک نویسه تغییر نمره این قسمت صفر خواهد بود.

۴-۱-۲-معیار F1

امتیاز F1 یک معیار متداول برای مسائل طبقه‌بندی است و به طور گسترده در QA استفاده می‌شود. موقعی مناسب

[1] Learning Rate
[2] epoch
[3] https://rajpurkar.github.io/SQuAD-explorer/

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.4.107]

است که ما به صحت[1] و بازخوانی[2] (حساسیت[3]) اهمیت بدهیم. در این مورد، بر روی کلمات پیش‌بینی‌شده در برابر مواردی که در پاسخ واقعی هستند، محاسبه می‌شود. تعداد کلمات مشترک بین پیش‌بینی و جواب‌های درست اساس نمره F1 است: صحت نسبت تعداد کلمات مشترک به کل کلمات پیش‌بینی‌شده و فراخوانی نسبت تعداد کلمات مشترک به تعداد کل کلمات در جواب‌های درست اصلی‌است.

$$\mathbf{F1} = \frac{\mathbf{2}}{\text{صحت}^{-1} + \text{حساسیت}^{-1}} = \mathbf{2}.\frac{\text{حساسیت} \times \text{صحت}}{\text{حساسیت} + \text{صحت}}$$

۴-۱-۳- اندازه فایل مدل‌ها

باتوجه‌به اینکه یکی از محدودیت‌های مدل‌های پردازش زبان طبیعی حجم بالای فایل‌های مدل آموزش‌دیده است، لذا در این بررسی اندازه فایل مدل نیز با اندازه مگابایت گزارش می‌شود تا اندازه مدل و دقت مدل‌ها آنها قابل‌اندازه‌گیری باشد و زمانی که محدودیت در استفاده از مدل‌های حجیم داریم، می‌شود با درصد کمی دقت پایین‌تر از این مدل‌ها استفاده کرد.

۴-۲- داده‌های ارزیابی

به‌منظور ارزیابی مدل‌ها ده درصد از سال‌ها به‌عنوان داده آزمون در نظر گرفته شدند که در نسخه ۱ دادگان FarsiQuAD برابر با هزار جفت پرسش‌وپاسخ می‌شود و در نسخه ۲ تعداد جفت سؤال‌وجواب‌ها برابر دوهزار است. در مدل‌هایی که داده‌های آموزشی با سایر مجموعه‌داده‌ها ترکیب شده بودند در موقع تست نیز داده‌های تست آن با هم ترکیب شده و نتایج گزارش شده‌است.

۴-۳- نتایج مدل‌های آموزش‌داده‌شده

باتوجه‌به اینکه مدل‌های آموزش‌دیده در چندین حالت مختلف بودند، لذا در ادامه به تفکیک نتایج آموزش مدل‌ها ارائه می‌شود. مقادیر ارزیابی استخراج‌شده برای سؤال‌های دارای پاسخ هستند.

۴-۳-۱- نتایج آموزش مدل‌ها بر روی داده‌های FarsiQuAD نسخه ۱

مدل‌های آموزش‌دیده‌شده بر اساس این دادگان (نسخه ۱) دارای نتایج یادشده است.

همان‌طور که در گزارش مشخص است، در این آموزش‌ها مدل پیش‌آموزش دیده‌شده با زبان فارسی نتایج بهتری را کسب کرده و تعداد مراحل ۳ با نرخ یادگیری 3e-5 نیز نتایج بهتری نسبت به بقیه هایپر پارامترها داشته‌اند.

(جدول - ۴). نتایج معیار F1 برای مدل های آموزش دیده با دادگان FarsiQuAD در نرخ‌های یادگیری و مراحل آموزش متفاوت

(Table -4) F1 Metric Results for Models Trained on FarsiQuAD Data at Different Learning Rates and Training Stages

نام مدل پیش آموزش دیده	η(3e-5) e(3)	η(3e-5) e(7)	η(5e-5) e(3)	η(5e-5) e(7)
bert-m-cased	84.92	83.73	84.83	83.71
bert-m-uncased	83.11	83.62	83.84	82.68
distilbert-uncased	53.99	57.64	47.85	56.10
xlm-roberta-base	81.48	81.78	81.35	80.01
bert-parsbert-uncased	84.94	84.93	83.09	82.43
bert-fa-uncased	85.38	84.88	85.10	84.35
bert-fa-zwnj	79.21	79.29	80.10	80.06
distilbert-fa-zwnj	78.95	77.77	77.83	77.67

(جدول - ۵). معیار تنظیم دقیق برای مدل‌های آموزش‌دیده با دادگان FarsiQuAD در نرخ‌های یادگیری و مراحل آموزش متفاوت

(Table -5). Precision Metric for Models Trained on FarsiQuAD Data at Different Learning Rates and Training Stages

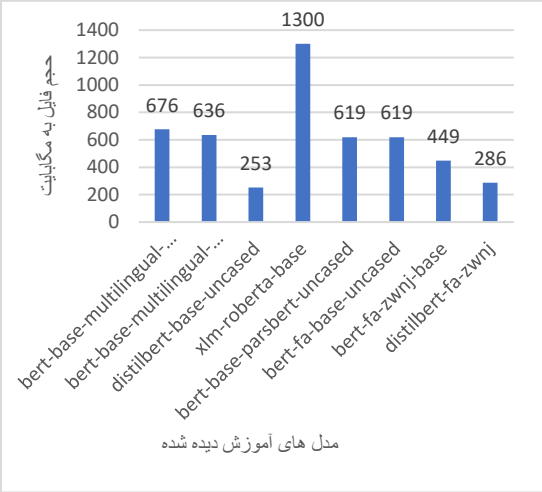
مدل های پیش آموزش دیده	η(3e-5) e(3)	η(3e-5) e(7)	η(5e-5) e(3)	η(5e-5) e(7)
bert-m-cased	76.95	76.55	75.53	75.63
bert-m-uncased	75.13	75.23	74.92	73.10
distilbert-uncased	38.98	43.76	32.79	40.10
xlm-roberta-base	72.18	72.69	70.86	70.05
bert-parsbert-uncased	76.55	76.24	73.60	73.30
bert-fa-base-uncased	77.16	76.14	76.04	75.33
bert-fa-zwnj-base	68.53	70.05	71.78	69.85
distilbert-fa-zwnj	68.93	67.82	66.70	67.41

به‌منظور مقایسه حجم فایل مدل‌های آموزش‌دیده شده در شکل (۱۰) اندازه حجم این مدل‌های آموزش‌دیده با دادگان FarsiQuAD به مگابایت گزارش شده‌اند.

[1] precision
[2] recall
[3] sensitivity

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.4.107]

با مقایسه حجم‌های مدل‌ها، مشخص است که مدل DistilBERT حجم فایل مدل خیلی کمتری نسبت به سایر مدل‌ها دارد؛ درحالی‌که معیار F1 و تنظیم دقیق برای این مدل اختلاف زیادی را ندارد.



(شکل – ۸). حجم فایل مدل های آموزش داده شده (MB) با دادگان FarsiQuAD

(Figure-8). Size of trained model files (MB) with the FarsiQuAD dataset

۴-۳-۲- نتایج آموزش مدل‌ها بر روی داده‌های دو دادگان زبان فارسی

بعد از ترکیب داده‌های تست دادگان سؤال‌وجواب FarsiQuAD و PersianQA نتایج به شرح شکل (۴-۲۱) حاصل شدند. این نتایج برای معیارهای دارای پاسخ‌است.

(جدول-۶) . نتایج مدل تجمیعی با داده تست دو دادگان زبان فارسی

(Table-6). Aggregate Model Results on Persian Language Test Data

مدل‌های پیش آموزش دیده‌شده	تطابق دقیق	F1
bert-m-uncased	64.12	80.51
distilbert-base-uncased	31.48	49.06
bert-parsbert-uncased	64.55	81.21
bert-fa-uncased	65.34	81.09
bert-fa-zwnj-base	59.78	75.81
distilbert-fa-zwnj	56.17	72.69
bert-base-m-cased-dutch-squad2	68.40	83.26

مدل چندزبانه که قبلاً با SQuAD2 آموزش‌دیده در این مرحله نتایج بهتری را نسبت به سایر مدل‌ها داده‌است.

بر اساس نتایج حاصل در جدول (۷) همچنان مدل DistilBERT دارای حجم بسیاری پایینی نسبت به سایر مدل‌ها بوده، ولی معیار F1 و تنظیم دقیق به‌طورتقریبی ده درصد کمتر را نسبت به سایر مدل‌ها ارائه داده‌است.

(جدول- ۷). حجم فایل مدل آموزش داده شده با دو دادگان و مقایسه با معیار F1

(Table-7). Size of Trained Model Files with Two Datasets and Comparison Based on F1 Metric

مدل های پایه پیش آموزش دیده	حجم فایل	معیار F1
bert-m-uncased	۶۳۶	۸۰.۵۱
distilbert-uncased	**۲۵۳**	۴۹.۰۶
bert-parsbert-uncased	۶۱۹	۸۱.۲۱
bert-fa-uncased	۶۱۹	۸۱.۰۹
bert-fa-zwnj-base	۴۴۹	۷۵.۸۱
distilbert-fa-zwnj	۲۸۶	۷۲.۶۹
bert-m-finetuned-dutch-squad2	۶۷۶	**۸۳.۲۶**

در ادامه مجدد تنها با داده‌های تست دادگان FarsiQuAD عملیات تست انجام شده تا بررسی شود آموزش با دو مجموعه دادگان چقدر می‌تواند نتایج حاصل از مجموعه تست داده‌های دادگان ایجادشده را افزایش دهد.

با بررسی این موضوع در جدول (۷)، آموزش مدل تجمیعی با دو دادگان باعث افزایش کمی در نتایج دادگان شده‌است و نشان‌دهنده تفاوت در سؤالات استخراج شده در دو مجموعه دادگان دارد.

در این قسمت نیز مدل چندزبانه گوگل که در قبل توسط دادگان SQuAD آموزش‌دیده نتایج بهتری را نسبت به سایر مدل‌ها ارائه کرده‌است.

(جدول-۸). نتایج مدل های تجمیعی بر روی داده های تست FarsiQuAD

(Table – ۸). Aggregate Model Results on FarsiQuAD Test Data

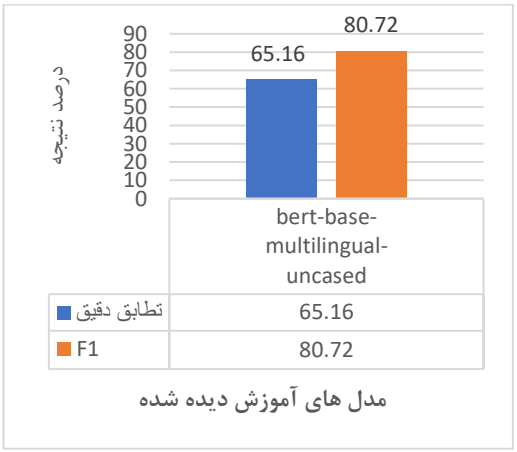
نام مدل‌های پیش آموزش دیده	تطابق دقیق	F1
bert-m-uncased	76.04	85.56
distilbert-uncased	39.09	55.86
bert-parsbert-uncased	76.24	86.15
bert-fa-uncased	76.24	86.15
bert-fa-zwnj-base	72.18	82.24
distilbert-fa-zwnj	68.02	79.47
bert-m-cased-finetuned-dutch-squad2	77.97	86.81

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

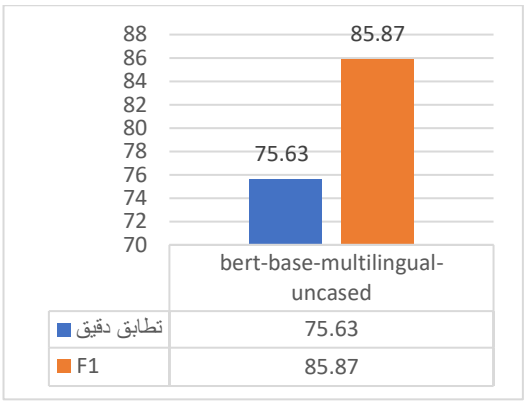
۴-۳-۴- نتایج آموزش مدل‌ها با سه مجموعه دادگان

با ارزیابی مدل آموزش‌داده‌شده برای سه مجموعه دادگان پرسش‌وپاسخ شامل (SQuADv2، دادگان FarsiQuAD، PersianQA) نتایج ذیل حاصل شدند.



(شکل -۹). نتایج آموزش مدل با سه دادگان پرسش و پاسخ و تست بر روی تجمیع دو دادگان فارسی

(Figure-9) Results of model training with three question-and-answer datasets and testing on the aggregation of two Persian datasets



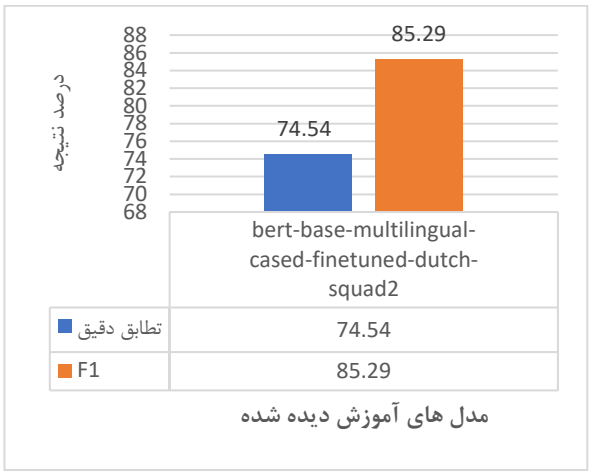
(شکل - ۱۰). نتایج آموزش مدل با سه دادگان پرسش و پاسخ و تست با داده‌های FarsiQuAD

(Figure -10). Results of model training with three question-and-answer datasets and testing with FarsiQuAD data

در بررسی به‌دست‌آمده مشخص شد آموزش مدل چندزبانه به‌صورت همزمان با سه مجموعه‌داده‌های عنوان‌شده به مقدار کمی نتایج را بهبود می ببخشد؛ ولی هنوز نتایج ضعیف‌تری نسبت به مدل‌های آموزش‌داده‌شده تک‌زبانه به زبان فارسی دارد.

۴-۳-۴- نتایج آموزش مدل‌ها با نسخه ۲ دادگان FarsiQuAD

با آموزش مدل با نسخه ۲ دادگان و تجمیع با مجموعه‌داده دادگان قبلی زبان فارسی، نتایج زیر برای مدل حاصل شد. مقادیر بر اساس نتایج دارای پاسخ ارائه شده‌است. بررسی‌ها نتایج خیلی خوب این مدل برای بررسی داده‌هایی که سیستم باید بداند جواب آن را نمی‌داند ارائه می‌دهد.



(شکل - ۱۱). نتیاج حاصل از آموزش مدل با نسخه ۲ دادگان FarsiQuAD

(Figure -11). Results obtained from model training with version 2 of the FarsiQuAD dataset

۴-۳-۵- مقایسه مدل ایجادشده با کارهای دیگر

باتوجه‌به اینکه مسئله پرسش‌وپاسخ جزو مسائل سخت طبقه‌بندی می‌شوند، لذا حل این موضوع با روش‌های عادی به‌سختی حل می‌شوند و بیشتر کارهای جدید از طریق مدل‌های یادگیری عمیق بوده که زبان را یاد گرفته‌اند، استفاده می‌شود؛ لذا در این قسمت داده‌های تستی با تعدادی از مدل‌های پیش آموزش‌دیده‌شده به زبان فارسی و سایر زبان‌ها بررسی شده و نتایج با بهترین مدل آموزش‌دیده‌شده در این مقاله مقایسه شده‌است.

(جدول - ۹). مقایسه بهترین مدل آموزش دیده‌شده با سایر مدل‌ها بر اساس داده‌های تست دادگان FarsiQuAD

(Table -9) . Comparison of the Best Trained Model with Other Models Based on FarsiQuAD Test Data

نام مدل ها	تطابق دقیق	F1
BERT-m(My Model)	77.97	86.81
BERT-base(FA)	43.00	69.00
BERT-m(EN)	50.61	68.20
XLM-RoBERTa-large(FA)	27.1	66.35
BERT(FA)	37.32	65.64
XLM-RoBERTa-large(EN)	29.41	60.38
RoBERTa-base(EN)	0.91	3.88
RoBERTa-large(EN)	0.91	3.29

نتایج مندرج در جدول (۹) نشان می‌دهد که مدل آموزش‌داده‌شده با داده‌های دادگان FarsiQuAD توانسته است معیار تطبیق دقیق را ۲۷.۳۶ و معیار F1 را ۱۷.۸۱

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.4.107]

درصد ارتقا دهد. همچنین در این بررسی مشخص شد مدل‌هایی که به‌صورت چندزبانه و یا زبان فارسی‌آموزش دیده‌اند، نتایج قابل قبولی دارند و مدل‌هایی که با زبان‌های تک‌زبانه غیر از فارسی آموزش دیده باشند، نتایج مناسبی ندارند.

۵- نتیجه‌گیری

با گسترش روزافزودن استفاده از مدل‌های یادگیری عمیق، این مدل‌ها در حال جایگزین‌شدن با روش‌های قبلی و یا ترکیب با این مدل‌های هستند. هم اکنون بیشتر وظایف زبان‌های طبیعی از طریق یادگیری عمیق حل می‌شوند. مدل‌های یادگیری عمیق می‌توانند مفهوم متن را در خود داشته باشند و جستجوی‌های مفهومی را به‌خوبی پاسخ دهند. از قابلیت‌های خوب مدل‌های یادگیری عمیق سرعت اجرای مناسب آن‌ها است.

در این پژوهش تلاش شد وظیفه پرسش‌وپاسخ به زبان فارسی با یادگیری عمیق را ارتقا دهد. بدین منظور دادگانی از مجموعه جفت پرسش‌وپاسخ‌ها به زبان فارسی تهیه و نام FarsiQuAD را برای آن انتخاب شد. این دادگان می‌تواند در آموزش مدل‌های یادگیری عمیق که به‌طورمعمول با محدودیت منابع آموزشی دست‌وپنجه نرم می‌کند، مفید واقع شود.

برای عملیات استخراج داده‌ها بیش از ۲۵۰۰ مقاله از مقاله‌های نمونه استفاده و از آنها سؤال‌هایی گوناگون در حوزه‌های مختلف شامل اشخاص، اماکن، حیوانات و ... با انواع مختلف سؤال شامل چرایی، چیستی، چگونگی و ... به تعداد بیش از ده‌هزار سؤال اقدام شد.

برای آموزش مدل‌ها از ساختار ترنسفورمر [۹،۱۰] و تعداد هشت مدل پیش آموزش‌دیده‌شده با زبان طبیعی شامل مدل‌های معروف (BERT [۱۱]، ParsBert [۱۲]، DistilBERT [۱۳]، RoBERTa[۱۴]، XML[۱۵] ، XML-RoBERTa [۱۶]) که با متون زیاد زبان فارسی و یا چندزبانه آموزش‌دیده شده‌اند، استفاده و عملیات تنظیم آنها با آموزش مجدد بر روی دادگان FarsiQuAD انجام شد.

در ادامه تنظیم مدل‌های یادگیری عمیق، این مدل‌ها با تجمیع دو دادگان زبان فارسی FarsiQuAD و PersianQA و سه مجموعه دادگان پرسش‌وپاسخ (FarsiQuAD، PersianQA، SQuAD انگلیسی) آموزش داده شدند.

به‌منظور مقاوم‌سازی شبکه برای مقابله با استخراج جواب‌های نامربوط، مدل منتخب مراحل گذشته با نسخه شماره ۲ دادگان FarsiQuAD آموزش داده شد و به نتایج خوبی رسید.

۶- مراجع 6-Refrence

[1]. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250, 2016.

[2]. Yuanjun Li, Yuzhu Zhang, Question Answering on SQuAD 2.0 Dataset, s. University, Editor, 2018.

[3]. d'Hoffschmidt, M., Belblidia, W., Brendlé, T., Heinrich, Q., & Vidal, M. FQuAD: French question answering dataset. arXiv preprint arXiv:2002.06071, 2020.

[4]. Möller, T., Risch, J., & Pietsch, M. Germanquad and germandpr: Improving non-english question answering and passage retrieval. arXiv preprint arXiv:2104.12741, 2021.

[5]. 임승영, 김명지, & 이주열. KorQuAD: 기계독해를 위한 한국어 질의응답 데이터셋. 한국정보과학회 학술발표논문집, 539-541, 2018.

[6]. 김영민, 임승영, 이현정, 박소윤, & 김명지. KorQuAD 2.0: 웹문서 기계독해를 위한 한국어 질의응답 데이터셋. 정보과학회논문지, 47(6), 577-586, 2020.

[7]. So, B., Byun, K., Kang, K., & Cho, S. Jaquad: Japanese question answering dataset for machine reading comprehension. arXiv preprint arXiv:2202.01764, 2022.

[8]. Ayoubi MY Sajjad & Davoodeh Persianqa: a dataset for persian question answering. https://github.com/SajjjadAyobi/PersianQA, 2021.

[9]. Mozafari, J., Fatemi, A., & Nematbakhsh, M. A. BAS: an answer selection method using BERT language model. arXiv preprint arXiv:1911.01528, 2019.

[10]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[11]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[12]. Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. Parsbert: Transformer-based model for persian language understanding. Neural Processing Letters, 53(6), 3831-3847, 2021.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]

[13]. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108, 2019.
[14]. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
[15]. Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. arXiv preprint arXiv:1901.07291.
[16]. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
[17]. Persian Wikipedia. Available from: https://github.com/miladfa7/Persian-Wikipedia-Dataset



جواد فروتن راد کارشناسی ارشد خود را در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه صنعتی مالک اشتر دریافت کرده و هم‌اکنون دانشجوی دکترای هوش مصنوعی است. ایشان علاقه‌مند به استفاده از فناوری‌های نوین نظیر هوش مصنوعی در سایر حوزه‌های مهندسی و پیوند آن با صنعت است. زمینه‌های کاری ایشان عبارت از پردازش زبان طبیعی، فناوری مالی، هوش مصنوعی و هوش تجاری در حوزه کاری است.
نشانی رایانامه ایشان عبارت است از :
Forutanrad@gmail.com



مریم حور علی کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات گرایش تجارت الکترونیک از دانشگاه علم و صنعت ایران و دکترای خود را در گرایش مهندسی فناوری اطلاعات از دانشگاه تربیت‌مدرس دریافت کرده‌است. ایشان در حال حاضر استادیار و عضو گروه علمی فناوری اطلاعات و فرماندهی کنترل و هوش مصنوعی دانشگاه صنعتی مالک‌اشتر است. زمینه‌های پژوهشی موردعلاقه ایشان عبارت‌اند از: پردازش زبان طبیعی، مهندسی فناوری اطلاعات و سامانه‌های فازی.
نشانی رایانامه ایشان عبارت است از:
mhourali@mut.ac.it

محمدعلی کیوان راد استادیار گروه هوش مصنوعی و عضو هیئت‌علمی دانشگاه صنعتی مالک‌اشتر است.
نشانی رایانامه ایشان عبارت است از:
keyvanrad@aut.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.4.107]