

# ML۴RE - یادگیری ماشین برای مهندسی نیازمندی‌ها

## چکیده

مقدمه: تحقیقات در زمینه یادگیری ماشین برای مهندسی نیازمندی‌ها (ML۴RE) به تدریج توجه بیشتری از سوی محققان و عملی‌کنندگان به خود جلب کرده است. اگرچه تحقیقات پیشگامانه پتانسیل استفاده از تکنیک‌های یادگیری ماشین برای بهبود فرآیندهای مهندسی نیازمندی‌ها را نشان داده‌اند، اما یک مرور نظام‌مند و جامع از ادبیات علمی که دیدگاه صنعتی را نیز در بر گیرد، در دانشگاه‌ها وجود ندارد. به‌ویژه، هیچ‌یک از مرورهای موجود در زمینه ML۴RE به ادبیات خاکستری که عمدتاً از منابع عملی‌کنندگان منشأ می‌گیرد و بازتاب‌دهنده مسائل و چالش‌های واقعی در عمل است، توجه نکرده‌اند.

هدف: در این مقاله، ما یک بررسی نظام‌مند از انتشارات علمی در زمینه ML۴RE انجام می‌دهیم و آن را با نظرات عملی‌کنندگان از Stack Overflow تکمیل می‌کنیم تا یک مرور جامع از ادبیات ارائه دهیم. هدف تحقیق ما ارائه یک دیدگاه جامع از پیشرفت‌های کنونی در تحقیقات ML۴RE، بیان سوالات و چالش‌های اصلی در عمل مهندسی نیازمندی‌ها، درک فاصله بین تحقیق و عمل، و ارائه بینش‌های خود درباره چگونگی توسعه عملی این حوزه دانشگاهی در آینده است.

روش: ما به صورت نظام‌مند ۲۰۷ مقاله علمی در زمینه ML۴RE از سال ۲۰۱۰ تا ۲۰۲۲ را بررسی کردیم و همچنین ۳۷۵ سوال مرتبط با مهندسی نیازمندی‌ها در Stack Overflow و پاسخ‌های مربوطه را تحلیل کردیم. تحلیل ما شامل روندها، فعالیت‌ها و وظایف متمرکز بر مهندسی نیازمندی‌ها، راه‌حل‌های به‌کاررفته و داده‌های مرتبط بود. در نهایت، یک تحلیل مشترک انجام دادیم و نتایج هر دو بخش را با هم مقایسه کردیم. نتایج: بر اساس نتایج آماری از ادبیات جمع‌آوری‌شده، ما یک نقشه راه علمی را خلاصه کرده و تفاوت‌ها را تحلیل کردیم و توصیه‌های پژوهشی ارائه دادیم. پیشنهادات ما شامل توسعه دستیاران هوشمند پاسخگویی به سوالات با استفاده از مدل‌های زبان بزرگ، ادغام یادگیری ماشین در ابزارهای صنعتی و ترویج همکاری بین دانشگاه و صنعت است.

نتیجه‌گیری: این مطالعه با ارائه یک دیدگاه جامع از ML۴RE، بیان تفاوت‌های بین تحقیق و عمل، و پیشنهاد راه‌حل‌های عملی برای پر کردن شکاف بین دانشگاه و صنعت، به پیشرفت این حوزه کمک می‌کند.

## ۱ مقدمه

مهندسی نیازمندی‌ها (RE) یک مرحله اساسی در مراحل اولیه مهندسی نرم‌افزار (SE) است. اگرچه پژوهشگران به طور مستمر در حال بررسی روش‌ها و تکنیک‌هایی برای تسهیل فرآیندهای نیازمندی هستند، اما کل فرآیند مهندسی نیازمندی‌ها همچنان نیاز به تلاش دستی زیادی دارد (مثلاً استخراج نیازمندی‌های ذی‌نفعان از طریق مصاحبه یا طبقه‌بندی نیازمندی‌ها بر اساس یک طبقه‌بندی خاص). دلیل اصلی این موضوع این است که فعالیت‌های RE معمولاً نیاز به دانش عمیق حوزه و مهارت‌های تحلیل پیشرفته دارند که به طور کامل قابل اتوماسیون نیست.

در سال‌های اخیر، توسعه سریع فناوری یادگیری ماشین (ML) با بهبود قدرت محاسباتی تحریک شده است. کاربردهای موفق ML در زمینه‌هایی مانند پردازش زبان طبیعی، شناسایی تصویر و داده‌کاوی فرصت‌هایی را برای استفاده از تکنیک‌های ML در زمینه RE فراهم کرده است. استفاده از فناوری ML در RE یک رویکرد هوشمندانه‌تر و کارآمدتر برای مدیریت داده‌های نیازمندی‌ها ارائه می‌دهد. به عنوان مثال، ML می‌تواند در طبقه‌بندی خودکار نیازمندی‌ها کمک کند زیرا می‌تواند اطلاعات بالقوه نیازمندی‌ها را خلاصه کند.

علاوه بر این، با توسعه سریع تکنیک‌های اطلاعاتی، کار و زندگی روزمره ما دیجیتالی می‌شوند. در نتیجه، داده‌های مرتبط با نیازمندی‌ها بیشتر و بیشتر دیجیتالی و به‌طور عمومی در دسترس قرار می‌گیرند، که پژوهش در زمینه یادگیری ماشین برای مهندسی نیازمندی‌ها (ML۴RE) را ترویج می‌کند. به عنوان مثال، بررسی‌های کاربران از برنامه‌های موبایلی به طور گسترده‌ای برای استخراج نیازمندی‌های کاربران مورد بررسی قرار گرفته‌اند. تحقیقات قبلی ML۴RE را مورد بررسی قرار داده‌اند. اقبال و همکاران [۲] یک بررسی برای به‌دست‌آوردن نمای کلی از چگونگی کمک تکنیک‌های ML به فعالیت‌های RE انجام دادند. علاوه بر این، زمانی و همکاران [۳] یک مطالعه نگاشت از کاربردهای ML در RE انجام دادند و ۶۵ مقاله را برای

ارزیابی اثربخشی ML در اتوماسیون وظایف RE تحلیل کردند. کارهای آن‌ها بر کل فرآیند RE متمرکز بود و نحوه تأثیرگذاری و تسهیل تکنیک‌های ML در مراحل مختلف را روشن کردند.

به علاوه، برخی تحقیقات به فعالیت‌ها یا وظایف خاص RE می‌پردازند. به عنوان مثال، لیم و همکاران [۴] رویکردهای پیشرفته فعلی برای استخراج نیازمندی‌های مبتنی بر داده از منابع داده پویا را بررسی کردند. ما متوجه شدیم که این مطالعات عمدتاً بر انتشارات علمی متمرکز بوده و از ادغام بینش‌های حاصل از منابع ادبیات خاکستری، مانند وبلاگ‌ها و انجمن‌های صنعتی غافل بوده‌اند.

بر خلاف انتشارات علمی که عمدتاً توسط پژوهشگران منتشر می‌شوند، ادبیات خاکستری به‌طور مداوم توسط عملی‌کنندگان تولید می‌شود و بر "وضعیت عمل" نور می‌تاباند [۵]. همان‌طور که در [۶] اشاره شده، ادغام ادبیات خاکستری در مرورهای نظام‌مند ادبیات می‌تواند فاصله بین پژوهش‌های علمی و عملی را پر کند و دیدگاه جامع‌تری از چالش‌ها و راه‌حل‌ها ارائه دهد.

اگرچه تعداد مرورهای نظام‌مند ادبیات که ادبیات خاکستری را در مطالعات SE در نظر گرفته‌اند در حال افزایش است [۷، ۸]، اما در RE به اندازه کافی رایج نیستند. برای پر کردن این شکاف در زمینه RE، این مقاله قصد دارد یک مرور نظام‌مند از ادبیات در زمینه ML۴RE انجام دهد که با بینش‌های حاصل از ادبیات خاکستری منابع شده از Overflow Stack تکمیل شود.

هدف این مرور ادبیات سه بخشی است. بخش سفید شامل مرور ۲۰۷ مقاله منتشر شده بین سال‌های ۲۰۱۰ تا ۲۰۲۲ است که به‌طور خاص بر ML۴RE متمرکز است. در همین حال، بخش خاکستری شامل تحلیل ۳۷۵ سوال و پاسخ‌های مربوطه جمع‌آوری شده از مباحث Overflow Stack درباره فعالیت‌های RE در همان دوره است. در نهایت، تحلیل مشترک ما شامل مقایسه نتایج این دو بخش برای تشخیص شباهت‌ها و تفاوت‌های آن‌ها است. ما روندها، فعالیت‌های RE، وظایف RE، راه‌حل‌ها و داده‌های موجود در ادبیات را تحلیل می‌کنیم.

نتایج تحقیق نشان می‌دهد که هر دو بخش به تحلیل RE و مستندسازی نیازمندی‌ها علاقه‌مند هستند. فراتر از شباهت‌ها، بخش سفید تمایل به تمرکز بر استخراج نیازمندی‌ها دارد، در حالی که بخش خاکستری بیشتر بر مدیریت نیازمندی‌ها تأکید دارد. بخش سفید استفاده از تکنولوژی‌های ML مانند CNN SVM و BERT را برجسته می‌کند. در مقابل، بخش خاکستری بیشتر بر ابزارهایی مانند Jira TFS، Microsoft و DOORS Rational IBM و تکنیک‌های ML مانند TF-IDF POS، LDA و شبکه‌های عصبی تکیه دارد. علاوه بر این، بخش خاکستری توجه ویژه‌ای به داستان کاربر و مورد استفاده دارد که در بخش سفید نسبتاً کمتر مورد بررسی قرار گرفته است.

بر اساس این یافته‌ها، ما یک نقشه راه علمی خلاصه کرده و تحلیل دقیقی از تفاوت‌های بین بخش سفید و خاکستری ارائه می‌دهیم. سپس پیشنهادات پژوهشی ارائه می‌دهیم، از جمله توسعه دستیاران هوشمند پاسخگویی به سوالات با استفاده از مدل‌های زبان بزرگ و ادغام یادگیری ماشین در ابزارهای صنعتی. همچنین، همکاری بیشتر بین دانشگاه و صنعت را برای درک عمیق‌تر مشکلات پژوهشی واقعی و داده‌ها تشویق می‌کنیم.

در خلاصه، این مقاله چهار کمک اصلی را ارائه می‌دهد. اولاً، یک نمای جامع از وضعیت فعلی پژوهش‌های ML۴RE ارائه می‌دهیم. دوماً، شرایط واقعی عملی‌کنندگان RE را از طریق ادبیات خاکستری حاصل از Overflow Stack بررسی می‌کنیم. سوماً، فاصله بین پژوهش و عمل در حوزه ML۴RE را به‌ویژه در زمینه‌هایی که کمتر مورد توجه پژوهشگران قرار گرفته‌اند، برجسته می‌کنیم. و در نهایت، برای پر کردن فاصله بین صنعت و دانشگاه، پیشنهادات پژوهشی عملی در ML۴RE ارائه می‌دهیم.

در بخش‌های باقی‌مانده این مقاله، کارهای مرتبط را در بخش ۲ ارائه می‌دهیم. بخش ۳ پروتکل تحقیق برای مرور نظام‌مند ادبیات ما را ارائه می‌دهد. در سه بخش بعدی، نتایج این بررسی و پاسخ به سوالات پژوهشی را ارائه می‌دهیم. بخش ۴ نتایج بخش سفید، بخش ۵ بر بخش خاکستری تمرکز می‌کند و بخش ۶ نتایج تحلیل مشترک را ارائه می‌دهد. بر اساس نتایج، در بخش ۷ به بحث پرداخته و چندین پیشنهاد ارائه می‌دهیم. بخش ۸ شامل تحلیل تهدیدات به اعتبار این بررسی است. در نهایت، مقاله را در بخش ۹ نتیجه‌گیری می‌کنیم.