



## پروژه درس علم داده‌ها در مهندسی نرم‌افزار

مدرس: دکتر عباس حیدرنوری

نیمسال اول ۹۹-۰۰

### ۱. مقدمه

هدف از پروژه درس علم داده این است که شما بتوانید از داده‌ها و اطلاعات موجود، برای حل چالش‌های مهندسی نرم‌افزار استفاده نمایید. بنابراین، شما در ابتدا یکی از چالش‌های موجود در مهندسی نرم‌افزار را شناسایی می‌کنید و سپس، با استفاده از مخازن داده مناسب و تکنیک‌هایی که در این درس فراگرفته‌اید، راه‌حلی برای آن ارائه می‌دهید و آن را پیاده‌سازی می‌کنید. به این منظور، پنج نوع مجموعه داده پیشنهاد شده و برای هر کدام چالش‌هایی که می‌تواند مورد توجه قرار گیرد، ذکر شده است. بنابراین، در گام اول هر گروه چالش موردنظر خود را انتخاب می‌کند. در گام دوم، مروری بر پژوهش‌های موجود در این زمینه انجام می‌دهید تا به این طریق با راهکارهای موجود آشنا شوید و راهکار کلی خود را پیشنهاد دهید. سپس، در گام سوم به پیاده‌سازی راهکار پیشنهادی می‌پردازید. بنابراین، در بخش بعدی مجموعه داده‌ها، چالش‌های پیشنهادی و نکاتی در مورد نحوه دسترسی به هر مجموعه داده ذکر شده است. سپس، در بخش سوم و چهارم به ترتیب نکات مورد توجه برای گام دوم و سوم پروژه بیان می‌شود. در نهایت، در بخش پنجم، زمانبندی تحویل گام‌های پروژه آورده شده است.

### ۲. مجموعه داده‌ها و چالش‌های پیشنهادی

همانطور که در جدول ۱ مشاهده می‌کنید، پنج مجموعه داده به همراه چالش‌های پیشنهادی برای هر کدام آورده شده است. برای دسترسی به هر مجموعه داده موارد زیر را در نظر بگیرید:

- **GitHub**: برای جمع‌آوری مجموعه داده مناسب می‌توانید از API های موجود استفاده کنید. همچنین، می‌توانید از مجموعه داده <sup>۱</sup>SEOSS استفاده کنید. این مجموعه داده در سال ۲۰۱۹ جمع‌آوری شده است و اطلاعات مربوط به نیازمندی‌ها، گزارش‌های باگ و تاریخچه کد را از Git و Jira دریافت کرده است.
- **StackOverflow**: برای جمع‌آوری مجموعه داده مناسب می‌توانید با استفاده از پلتفرم ابری گوگل <sup>۲</sup> یا سایت Stack Exchange Data Explorer<sup>۳</sup> به صورت برخط روی داده‌ها کوئری بزنید و زیرمجموعه مورد نظر خود را دریافت کنید. برای اطلاعات بیشتر پیرامون ساختار پایگاه داده می‌توانید به این [لینک](#) مراجعه کنید.
- **TravisTorrent**: لینک دریافت مجموعه داده مورد نظر در صفحه MSR Challenge 2017 موجود می‌باشد.
- **SStuBs**: لینک دریافت مجموعه داده مورد نظر در صفحه MSR Challenge 2021 موجود می‌باشد.
- **Jira**: برای دریافت مجموعه داده مورد نیاز می‌توانید به پروژه <sup>۴</sup>jira-social-repository در GitHub مراجعه کنید.

<sup>۱</sup> <https://bit.ly/2wukCHc>

<sup>۲</sup> <https://cloud.google.com/bigquery/docs/bigquery-web-ui>

<sup>۳</sup> <https://data.stackexchange.com/stackoverflow/queries>

<sup>۴</sup> <https://github.com/marcoortu/jira-social-repository>

جدول ۱ مجموعه داده‌ها و چالش‌های پیشنهادی

ردیف	مجموعه داده	چالش‌های پیشنهادی
۱	StackOverflow	یافتن خبره
		پیش‌بینی زمان پاسخ
		برای آشنایی با چالش‌های بیشتر، مراجعه به MSR Challenge 2013 <sup>۱</sup>
		MSR Challenge 2015 <sup>۲</sup>
۲	GitHub	توصیه نوع بازآوری کد با استفاده از درخواست‌های ویژگی
		بررسی ارتباط بین باگ‌های گزارش‌شده و نوع بازآوری کد
		بررسی تأثیر بازآوری کد روی تعدادی از Code smell ها <sup>۳</sup>
		در تمام یا بخشی از یک پروژه
		اختصاص باگ به بهترین توسعه‌دهنده برای رفع آن
۳	TravisTorrent	ارائه رویکردی برای پیش‌بینی باگ
		برای آشنایی با چالش‌های موجود، مراجعه به MSR Challenge 2017 <sup>۴</sup>
۴	SStuBs	برای آشنایی با چالش‌های موجود، مراجعه به MSR Challenge 2021 <sup>۵</sup>
۵	Jira	ارتباط بین توسعه‌دهندگان و تأثیر آن روی فرایند توسعه

همچنین، در پروژه‌های حوزه بازآرایی کد، گروه‌ها می‌توانند از ابزار RefactoringMiner<sup>۶</sup> برای شناسایی انواع بازآرایی در کد استفاده کنند.

### ۳. نکات مورد توجه در گام دوم پروژه

لطفاً موارد زیر را در نظر بگیرید:

۱. انتظار داریم این گام یک دید جامعی از پژوهش‌های انجام‌شده در حوزه شما برایتان فراهم کند. بنابراین، سعی کنید پوشش خوبی از انواع پژوهش‌های این حوزه داشته باشید.
۲. در انتهای گزارش این گام باید با توجه به چالش انتخابی خود، راه‌حلی را مطرح نمایید و در گام بعدی روی آن متمرکز شوید. **توجه شود** که راه‌حل شما نباید صرفاً پیاده‌سازی یکی از رویکردهای موجود باشد و باید ایده‌های جدیدی مطرح شود.

<sup>۱</sup> <http://2013.msrconf.org/challenge.php>

<sup>۲</sup> <http://2015.msrconf.org/challenge.php>

<sup>۳</sup> شناسایی code smell ها می‌تواند به صورت rule based انجام شود.

<sup>۴</sup> <http://2017.msrconf.org/#/challenge>

<sup>۵</sup> <https://2021.msrconf.org/track/msr-2021-mining-challenge#Call-for-Mining-Challenge-Papers>

<sup>۶</sup> <https://github.com/tsantalis/RefactoringMiner>

۳. از آن جا که در حوزه های انتخابی شما مقالات زیادی منتشر شده است، تا حد امکان تمرکز خود را بر روی مقالاتی با رنک A و A\* قرار دهید. همچنین، لازم به ذکر است که توجه به مقالات سال های اخیر اهمیت زیادی دارد، اما این مسئله نباید سبب شود کارهای مهم و تأثیرگذار که لزوماً در این اواخر به چاپ نرسیدند از قلم بیفتند. سعی کنید دید جامعی از کارهای مرتبط داشته باشید.

نکته: برای سرچ رنک کنفرانس می توانید از <http://portal.core.edu.au/conf-ranks/> و برای ژورنال از <http://portal.core.edu.au/jnl-ranks/> استفاده کنید. البته، برای ژورنال ها بهتر است، اسم ژورنال را با sjr (Empirical Software Engineering sjr) در گوگل سرچ کنید. وارد سایت SCImago شوید. ژورنال های Q1 و Q2 کیفیت خوبی دارند. کنفرانس های رنک A و B معتبرند.

۴. **توجه شود** که انتظار می رود برای این سروی، شما تنها به خواندن بخش های Abstract و Conclusion هر مقاله اکتفا نکنید. از طرفی، نیازی هم نیست هر مقاله را با جزئیات کامل مورد مطالعه قرار دهید. به هر حال، برای هر مقاله باید هدف مقاله، رویکرد مورد استفاده به صورت کلی (حداقل باید نام الگوریتم های مورد استفاده و روند رویکرد ذکر شود) و خلاصه ای از نحوه ارزیابی آن شرح دهید.

۵. **توجه شود** که در سروی خود تنها به بیان نظر نویسنده ها اکتفا نکنید. سعی کنید دیدگاه خودتان را هم به گونه ای منعکس کنید. به عنوان مثال، باید در انتهای هر بخش مقایسه مختصری از پژوهش های مطرح شده در آن داشته باشید. این موضوع از آن جهت اهمیت دارد که به شما در گام بعدی کمک می کند به مسئله مورد نظر تسلط داشته باشید و رویکرد مناسبی را بتوانید ارائه دهید.

۶. اگر در حوزه شما سروی وجود دارد، می توانید از آن بهره ببرید ولی دقت کنید که تنها کار خود را محدود به آن نکنید و مقالات جدید را نیز پوشش دهید.

۷. تعداد مقالات اهمیت زیادی ندارد و مهم کیفیت کار شماست ولی خیلی کم هم نباید باشد. به عنوان مثال، برای یک گروه ۲ نفره حداقل ۱۵ مقاله باید مورد بررسی قرار گیرد.

۸. **توجه شود** که باید در انتهای سروی یا در انتهای هر بخشی از آن، پژوهش های مطرح شده به صورت خلاصه در جدولی آورده می شود. لازم نیست معیارهای زیادی داشته باشد ولی یک دید کلی برایتان فراهم می کند و در گام به شما کمک می کند.

۹. **توجه شود** که این گزارش باید در فرمت مقالات انجمن علمی کامپیوتر ایران ارائه گردد و شامل بخش های چکیده، کلمات کلیدی، مقدمه، کارهای پیشین، تعریف مسئله، نتیجه گیری و منابع باشد.

۱۰. در این گام بررسی مقالات معتبر، کیفیت گزارش پایانی (شامل رعایت ساختار مناسب، نکات نگارشی، ارجاع دادن در موارد لازم، مرتب بودن منابع، استفاده مناسب از جداول و اشکال) و مناسب بودن محتوای هر بخش حائز اهمیت است. نکته: مرتب بودن منابع در هر دو فاز مهم است. توجه شود که لزوماً رفرنسی که از گوگل دریافت می کنید، مرتب نیست. مرتب بودن به این معنا است که در تمامی رفرنس ها، ترتیب ذکر مشخصات یکسان باشد یا نوع اطلاعاتی که برای کنفرانس ها، ژورنال ها، کتاب ها یا غیره ارائه می شود، یکسان باشد. مثلاً در تمامی رفرنس ها مخفف نام یا نام کامل کنفرانس یا ژورنال قرار داده شود.

۱۱. برای این گام یک فایل pdf با نام DSSE\_2\_#groupnumber آپلود نمایید (مثلاً برای گروه ۳ : DSSE\_2\_3). شماره گروه ها اعلام خواهد شد.

#### ۴. نکات مورد توجه در گام سوم پروژه

لطفاً موارد زیر را در نظر بگیرید:

۱. در گام سوم شما باید راهکار پیشنهادی خود را پیاده‌سازی کنید. همچنین، باید راهکار خود را با روش‌ها و معیارهای مناسب حوزه خود مورد ارزیابی قرار دهید. تحلیل نتایج ارزیابی نیز از اهمیت بالایی برخوردار است. باید با تحلیل مناسب، ضعف‌ها و قوت‌هایی که در ارزیابی مشخص می‌شود را تحلیل کنید. به‌علاوه، استفاده از معیارهای مناسب نیز هائز اهمیت است.
۲. توجه شود که این گزارش باید در فرمت مقالات انجمن علمی کامپیوتر ایران ارائه گردد و بخش‌های این گام شامل (چکیده، کلمات کلیدی، مقدمه، پیاده‌سازی رویکرد پیشنهادی، ارزیابی و تحلیل رویکرد پیشنهادی (براساس معیارهای معتبر)، تهدیدات علیه اعتبار، نتیجه‌گیری و منابع) می‌باشد.
۳. کیفیت گزارش پایانی (شامل رعایت ساختار مناسب، نکات نگارشی، ارجاع دادن در موارد لازم، مرتب بودن منابع، استفاده مناسب از جداول و اشکال) و مناسب بودن محتوای هر بخش حائز اهمیت است.
۴. برای این گام یک فایل zip با نام DSSE\_3\_#groupnumber آپلود نمایید (مثلاً برای گروه ۳: DSSE\_3\_3). که در آن یک pdf با نام DS\_3\_#groupnumber قرار دارد که گزارش شما است و در یک فولدر به نام Implementation کدهای پیاده‌سازی شما قرار می‌گیرد. شماره گروه‌ها اعلام خواهد شد.

#### ۵. زمانبندی گام‌های پروژه

در جدول ۲ زمانبندی گام‌های پروژه آورده شده است. همچنین، سوالات خود را می‌توانید در پیازا مطرح کنید یا از طریق ایمیل [yasaman.abedini.ta@gmail.com](mailto:yasaman.abedini.ta@gmail.com) با بنده در تماس باشید. همچنین، لطفاً سوالات خود را در CW مطرح نکنید و از سایر ایمیل‌های بنده نیز استفاده نکنید.

جدول ۲ زمانبندی گام‌های پروژه

ردیف	گام‌های پروژه	مهلت تحویل
۱	گام ۱: انتخاب چالش	۱۳۹۹/۹/۳۰
۲	گام ۲: مقاله مروری	۱۳۹۹/۱۰/۲۶
۳	گام ۳: پیاده‌سازی راهکار پیشنهادی	۱۳۹۹/۱۱/۱۵

موفق باشید،

عابدینی

([yasaman.abedini.ta@gmail.com](mailto:yasaman.abedini.ta@gmail.com))