فرآیند پیاده‌سازی مدل بخشی پیچیده از چرخه عمر پروژه‌های یادگیری ماشین است [۶۵، ۱۰۶]. استفاده از روش‌شناسی DevOps [۹۷، ۱۱، ۳۲] و ابزارهایی برای یکپارچگی مداوم و ارائه مداوم (CI/CD) در توسعه نرم‌افزار [۳۴، ۲۹]، موفقیت‌های ثابتی داشته است.

یادگیری ماشین اغلب بخش کوچکی از سیستم‌های نرم‌افزاری است، اما یک نرم‌افزار که شامل یادگیری ماشین است، در مسیر CI/CD پیشروی نکرده است. سیستم‌های یادگیری ماشین پیچیدگی و مشکلات منحصر به فردی را به لوله‌های CI/CD اضافه می‌کنند [۹۱]. گسترش روش‌شناسی DevOps برای شامل شدن ویژگی‌های سیستم‌های یادگیری ماشین به طور گسترده‌ای به عنوان MLOps شناخته می‌شود [۶۹].

محصولات محبوب فعلی برای MLOps بیشتر متعلق به ارائه‌دهندگان ابری مختلف مانند آمازون، گوگل و مایکروسافت هستند [۲، ۱، ۱۰]. راه‌حل‌های امتیازی کارآمدی برای ادامه‌پذیری و شفافیت در لوله‌های کاری نرم‌افزاری ارائه می‌دهند، اما در عین حال منجر به بسته‌بندی شدید با فروشنده می‌شوند. خدمات ارائه‌دهندگان ابری اغلب راه‌حل قابل قبولی برای شرکت‌هایی که در دستگاه‌های نرم‌افزاری تنظیمات نظارتی یا مشکلات حریم خصوصی کاربر دارند، نیستند. آن‌ها به راه‌حل MLOps نیاز دارند که قابل اجرا در ابرها و دستگاه‌های داخلی باشد. راه‌حل‌های فعلی منبع‌باز و غیرابرمنبعی که برای لوله‌های MLOps از ابتدا تا انتها معمولاً ناقص، ناکامل، عمومی نشده، سخت یادگیری و استفاده یا نامناسب برای تولید هستند.

هدف این پایان‌نامه، طراحی یک لوله MLOps برای برنامه‌های بومی ابری است که نمی‌توانند به نرم‌افزار امتیازی یا شرایط ارائه‌دهنده ابری اتکا کنند. ما یک لوله MLOps منبع‌باز و ساده برای استفاده ارائه می‌دهیم که محیطی بر پایه Kubernetes را برای CI/CD و نظارت بر سیستم‌های یادگیری ماشین فراهم می‌کند. استنتاج‌های مدل تولید شده در لوله قابل ردیابی به داده‌ها و پارامترهای خود برای اهداف نظارتی هستند. همچنین یک رابط برای بررسی مدل با ردیابی خودکار داده‌ها، پارامترها و معیارها فراهم می‌کند.

در فصل ۲، رویکرد تحقیق، سوالات تحقیق و اهداف طراحی برای لوله MLOps را معرفی می‌کنیم. در فصل ۳، روش‌شناسی DevOps را تعریف می‌کنیم، لوله‌های CI/CD را توصیف می‌کنیم و نیازمندی‌های آن‌ها را توضیح می‌دهیم، دلیل اینکه DevOps و CI/CD در توسعه نرم‌افزار مفید است را بیان می‌کنیم و مراحل اجرای آن را تعریف می‌کنیم. فناوری‌های بومی ابری را شرح می‌دهیم تا استفاده از آن‌ها را درک و توجیه کنیم. در فصل ۴، روش‌های DevOps و لوله‌های CI/CD را به سیستم‌های یادگیری ماشین گسترش و مقایسه می‌کنیم تا مشکلات و نیازهای یک لوله MLOps را پیدا کنیم. فصل ۵ به جزئیات طراحی و اجرای پیشنهاد ما می‌پردازد، از جمله بررسیی بر ابزارها، روش‌ها و معماری لوله. توصیف می‌کنیم که چگونه اجزای و مراحل DevOps را که در بخش‌های قبلی تعریف شده‌اند، پیاده‌سازی می‌کند و نیازهای لوله‌های CI/CD سیستم‌های یادگیری ماشین را برآورده می‌کند. فصل ۶ به بررسی پیاده‌سازی نیازهای MLOps یک پروژه ساده یادگیری ماشین در این لوله MLOps می‌پردازد. فصل ۷ شامل بحث در مورد محدودیت‌ها، آینده و کارهای مرتبط است. در فصل ۸ به پایان‌نامه با استنتاج‌ها خاتمه می‌دهیم.

عنوان: MLOps: تسهیل عملیات یادگیری ماشین

مقدمه: MLOps، که مخفف عملیات یادگیری ماشین است، یک رشته است که هدف آن بهبود کارآیی، قابلیت اطمینان و مقیاس پذیری جریان کارهای یادگیری ماشین در طول دوره عمر کامل آنها است. با افزایش انتشار مدل‌های یادگیری ماشین برای تصمیم‌گیری‌های مبتنی بر داده، MLOps به عنوان یک شاخصه حیاتی برای پر کردن شکاف بین توسعه، راه‌اندازی و بهبود مداوم سیستم‌های یادگیری ماشین ظاهر می‌شود. این صفحه تعریف جامعی از MLOps ارائه می‌دهد و برجسته‌ترین اجزا و مزایای آن را برجسته می‌کند.

تعریف MLOps: MLOps شامل مجموعه‌ای از روش‌ها، ابزارها و رویکردها است که همکاری، اتوماسیون و نظارت را در چرخه عمر پروژه‌های یادگیری ماشین تسهیل می‌کند. این رشته از عناصری چون DevOps، مهندسی داده و مهندسی نرم‌افزار سنتی استفاده می‌کند تا چالش‌های منحصر به فرد مرتبط با توسعه و راه‌اندازی مدل‌های یادگیری ماشین را برطرف کند.

اجزای کلیدی MLOps:

ادغام و تحویل مداوم (CI/CD): MLOps بر استفاده از لوله‌های CI/CD برای اتوماسیون فرآیند یکپارچه‌سازی تغییرات کدها، آزمون مدل‌های یادگیری ماشین و راه‌اندازی آنها در محیط‌های تولید تأکید می‌کند. این امر به این معناست که سیستم‌های یادگیری ماشین به صورت سریع و قابل اعتماد توسعه، اعتبارسنجی و عرضه می‌شوند.

کنترل نسخه: سیستم‌های کنترل نسخه، مانند گیت، نقش حیاتی در MLOps را با امکان ردیابی تغییرات کدها، مجموعه داده‌ها و فایل‌های مدل دارند. آنها همکاری بین اعضای تیم را تسهیل می‌کنند، سابقه‌ای از نسخه‌های مدل ایجاد می‌کنند و قابلیت تکرارپذیری را فراهم می‌کنند.

هماهنگی زیرساخت: MLOps شامل مدیریت زیرساخت مورد نیاز برای بارهای کاری یادگیری ماشین است. ابزارهای هماهنگی زیرساخت مانند Kubernetes به مقیاس‌پذیری کارآمد، تخصیص منابع و ظرفیت کانتینریزه کردن مدل‌های یادگیری ماشین کمک می‌کنند و راه‌اندازی قابل اعتماد و مقیاس‌پذیر را تضمین می‌کنند.

نظارت و مدیریت مدل: نظارت بر عملکرد و رفتار مدل‌های یادگیری ماشین در محیط‌های تولیدی جزء جنبه‌های حیاتی MLOps است. این شامل پیگیری معیارهای کلیدی، تشخیص ناهنجاری‌ها و امکان مداخله پیشگیرانه برای حفظ دقت و قابلیت اعتماد مدل است. مدیریت مدل شامل نسخه‌بندی، کاتالوگ‌سازی و راه‌اندازی به‌روزرسانی‌های مدل در محیط‌های تولیدی است.

مدیریت داده و ویژگی: MLOps اهمیت مدیریت صحیح داده‌ها و ویژگی‌ها را در جریان کارهای یادگیری ماشین تأیید می‌کند. این شامل پیش‌پردازش داده‌ها، نسخه‌بندی داده‌ها، خطوط لوله ویژگی و تضمین کیفیت داده است. مدیریت صحیح داده و ویژگی، تکرارپذیری و قابلیت اعتماد مدل‌های یادگیری ماشین را بهبود می‌بخشد.

مزایای MLOps:

افزایش بهره‌وری: MLOps وظایف تکراری را اتوماسیون کرده و زحمت دستی را کاهش می‌دهد، که این امر به محققان داده و مهندسان امکان می‌دهد تا بر روی وظایف اصلی یادگیری ماشین تمرکز کنند. این روند توسعه و راه‌اندازی را سریعتر می‌کند و بهبود بهره‌وری را دارد.

همکاری بهبود یافته: MLOps همکاری بین محققان داده، مهندسان و سایر ذینفعان در پروژه‌های یادگیری ماشین را ترویج می‌دهد. با فراهم کردن یک محیط مشترک، کنترل نسخه و جریان کارهای استاندارد، MLOps همکاری مؤثر و به اشتراک‌گذاری دانش را تسهیل می‌کند.

بهبود قابلیت اعتماد مدل: MLOps آزمون، اعتبارسنجی و مکانیزم‌های نظارت قوی را دربرمی‌گیرد، تضمین کننده قابلیت اعتماد و پایداری مدل‌های یادگیری ماشین در محیط‌های تولیدی است. با نظارت بر عملکرد مدل و رفع مشکلات به موقع، MLOps کمک به جلوگیری از خطاها و انحراف داده‌ها می‌کند.

قابلیت مقیاس‌پذیری و تکرارپذیری: MLOps با بهره‌گیری از قابلیت مقیاس‌پذیری کانتینریزه کردن و هماهنگی زیرساخت، مقیاس‌پذیری بی‌دردسر جریان کارهای یادگیری ماشین را فراهم می‌کند. این امر اطمینان می‌دهد که سیستم‌های یادگیری ماشین قادر به کار با مجموعه داده‌های بزرگ، تقاضای کاربران بیشتر و نیازهای محاسباتی پیچیده باشند. علاوه بر این، MLOps تکرارپذیری را با نگهداری سابقه و نسخه‌بندی مدل‌ها، داده‌ها و وابستگی‌های مرتبط فراهم می‌کند.

نتیجه‌گیری: MLOps نقش حیاتی در عملیاتی کردن جریان کارهای یادگیری ماشین دارد و سازمان‌ها را قادر می‌سازد تا به طور کارآمد مدل‌های یادگیری ماشین را توسعه، راه‌اندازی و حفظ کنند. با ترکیب اصول DevOps، مهندسی داده و مهندسی نرم‌افزار، MLOps همکاری، اتوماسیون فرایندها و تضمین قابلیت اعتماد و مقیاس‌پذیری سیستم‌های یادگیری ماشین را فراهم می‌کند. پذیرفتن MLOps سازمان‌ها را قادر می‌سازد تا توانمندی کامل اقدامات یادگیری ماشین خود را فعال سازند و به نتایج کسب و کار موثر دست یابند.