



باسمه تعالی

برگه تعریف موضوع پایان نامه کارشناسی ارشد

شماره دانشجویی: ۴۰۰۲۰۵۶۴۱

تعداد واحد: ۶

شماره درس پایان نامه: ۲۵۸۱۰

نام و نام خانوادگی: ابوالفضل یاریان کوپائی

دانشکده برق: گرایش سیستم های

الکترونیک دیجیتال

نام استادراهنمای پروژه: دکتر متین هاشمی

نام استادراهنمای همکار(درصد همکاری نام استاد مشاور: -

عنوان کامل پروژه:

فارسی: طراحی و پیاده سازی پلتفرم MLOps مبتنی بر GPU به صورت ابری

لاتین: Design and Implementation of GPU-based MLOps Cloud Platform

۳-خودکارسازی

۲-یادگیری ماشین

کلمات کلیدی: ۱- MLOps

نتایج تحقیقات و پایان نامه های دانشجویان و درآمدهای مالی ناشی از آن متعلق به دانشگاه صنعتی شریف بوده و هیچ شخصیت حقیقی یا حقوقی بدون اجازه دانشگاه حق بهره برداری از آن را نخواهد داشت. همچنین کلیه حقوق اعم از چاپ و تکثیر و نسخه برداری، ترجمه و اقتباس، و نظایر آن در محیط های مختلف از پایان نامه برای دانشگاه صنعتی شریف محفوظ است. نقل مطالب با ذکر مأخذ بلامانع است.

موضوع مورد تایید و اتمام به موقع پایان نامه در چهارچوب آیین نامه های مصوب دوره های تحصیلات تکمیلی، امکانات موجود در دانشکده و اعتبارات مربوط به خرید اقلام مصرفی جزئی امکان پذیر است.

امضای استادراهنمای پایان نامه:

امضای دانشجو:

نظر گروه:

مورد تصویب قرار گرفت.

موضوع در جلسه مورخ

نام و امضای مدیر گروه:

نظر شورای تحصیلات تکمیلی دانشکده:

مورد تصویب قرار گرفت.

موضوع در جلسه مورخ

توضیحات:

نام و امضای معاون تحصیلات تکمیلی:

شرح مختصر موضوع پروژه :

MLOps مجموعه ای از متدها در حوزه یادگیری ماشین برای یکپارچه سازی توسعه (Dev) سیستم ML و عملیات (Ops) آن می باشد. به عبارت دیگر تلاش برای خودکار سازی و مانیتور تمامی گام های ساخت سیستم های ML، یکپارچه- سازی، تست، انتشار، استقرار و مدیریت زیرساخت آنها است. یکی از عوامل مهمی که موجب مطرح شدن MLOps در فناوری های جدید شده است، پیچیدگی ها و چالش های حول یک کد یادگیری ماشین است که تنها بخش کوچکی از سیستم های واقعی یادگیری ماشین هستند.[4]

واحدهای پردازش ابری گرافیکی (**GPU Cloud**) وظیفه ارائه منابع لازم برای سرعت بخشی به یک برنامه را دارند. این عملکرد بدون نیاز به داشتن یک پردازنده گرافیکی در اختیار کاربر است. به طوری که پردازش گرافیکی مورد نیاز کاربر در فضای ابری قرار می گیرد.

سابقه انجام کارهای قبلی: [1][2][3]

در سال های اخیر ابزار ها متن باز مختلفی برای کمک به خودکار سازی فرایند ها یادگیری ماشین پدید آمده است. توجه داشته باشید که ابزار های مختلف مراحل مختلفی را در این فرایند خودکار سازی می کنند که معمولاً نیز به زبان های پایتون و R توسعه یافته اند. در زیر به برخی از آن ها اشاره می کنیم :

۱- **ابزارهای پیش پردازش داده :** به دو دسته اصلی ، برچسب گذاری داده ها (data labeling) و مدیریت نسخه داده (data versioning) تقسیم می شوند. ابزار های برچسب گذاری داده وظیفه برچسب گذاری داده های حجیم مثل تصاویر یا صدا ، حاشیه نویسی و ... دارند که براساس دقت برچسب گذاری آن ها، مدل ها عملکرد متفاوتی خواهند داشت . از جمله این ابزارها می توان به Labelbox ، tagtog اشاره نمود. مهم ترین ابزارهای این بخش، مدیریت نسخه داده می باشد. این ابزارها زمانی به کار می آیند که دیتاست مورد بررسی، دچار تغییر شود (تصحیح – افزایش – کاهش). نسخه سازی برای ردیابی تغییرات مرتبط با داده می تواند زمان پردازش ها را کاهش داده و به محقق برای افزایش سرعت آموزش یا ارزیابی کمک کند. از مهم ترین ابزار های مدیریت نسخه داده، DVC (Data version control) می باشد. محققین یادگیری ماشین می توانند با استفاده از آن به مدیریت مدل های خود از یکسو و هم چنین اشتراک گذاری آن ها از سوی دیگر بپردازند. DVC شباهت زیادی از بعد معماری و ساختار با سیستم مدیریت نسخه GIT دارد. از ابزارهای مهم دیگر در این بخش می توان به Pachyderm ، MLflow و IBM Watson ML نیز نام برد.

۲- **ابزارهای مدل سازی :** به چند دسته متفاوت تقسیم می شوند.

(۱) **Feature Engineering :** ابزارهایی که با آن ها ویژگی ها را از یک مجموعه داده خام استخراج می کنیم تا مجموعه داده های آموزش بهینه را ایجاد کنیم که از مهم ترین ابزارها می توان به Featuretools ، AutoFeat و TSFresh اشاره کرد.

(۲) **Experiment Tracking :** برای نظارت بر نسخه داده های هر آزمایش و نتایج آن و هم چنین مقایسه بین آزمایش های مختلف استفاده می شود. این ابزارها تمام اطلاعات لازم در مورد آزمایش ها مختلف را ذخیره می کنند زیرا توسعه پروژه های یادگیری ماشین شامل اجرای آزمایش های متعدد با مدل های مختلف است. از مهم ترین ابزار های مورد استفاده برای این منظور می توان به Comet ، Weight and Biases و Neptune.ai اشاره کرد .

(۳) **Hyperparameter Optimization :** فرآیند جستجو و انتخاب هایپرپارامترهایی (مانند اندازه شبکه عصبی یا regularization term در رگرسیون) که عملکرد مطلوبی را برای مدل های یادگیری ماشین ارائه می دهد را

خودکار سازی می کند. از مهم ترین ابزار های مورد استفاده برای این منظور می توان به Optuna، Hyperopt و Scikti-Optimize اشاره کرد.

۳- ابزارهای عملیاتی سازی : خود شامل ابزارهای استقرار (deployment tools) مانند Algorithmia و Kubeflow و ابزارهای نظارت (monitoring) می باشند. نظارت بر مدل های یادگیری ماشین یکی از جنبه های کلیدی هر پروژه یادگیری ماشین است زیرا عملکرد مدل یادگیری ماشین پس از استقرار مدل به دلیل تغییرات در داده های ورودی در طول زمان بدتر خواهد شد. به این تغییر data drift یا concept drift می گویند. ابزارهای نظارت مدل، data drift را در طول زمان تشخیص می دهند و امکان تنظیم هشدارها در صورت بروز مشکلات عملکردی را فراهم می کنند. از مهم ترین ابزارهای نظارت نیز می توان به Evidently AI ، Fiddler و Arize اشاره کرد.

البته ابزارهایی هم نظیر Google Cloud Platform و Microsoft Azure هستند که تمامی سیکل یک برنامه یادگیری ماشین (end-to-end) را به منظور خودکار سازی پوشش می دهند.

اهداف مشخص انجام پروژه:

یادگیری ماشین طی سالیان اخیر با دسترسی بیشتر به داده ها، پیشرفت در قابلیت محاسبات و راه حل های ذخیره سازی به سرعت در حال رشد و ارائه امکانات جدید می باشد. با توجه به لزوم استفاده از GPU به عنوان هسته های پردازشی به منظور افزایش سرعت در آموزش شبکه های عصبی عمیق ، رندرینگ تصویر و پردازش کلان داده، امروزه نیاز است که این روند یادگیری مدل ها به صورت خودکار انجام شده و هم چنین این پردازش ها به صورت ابری باشد. هدف ما طراحی و پیاده سازی یک پلتفرم MLOps به منظور خودکارسازی فرایند های یادگیری ماشین بر روی GPU و به صورت ابری می باشد.

مراحل انجام کار :

- ۱- مطالعه و بررسی مفاهیم اصلی MLOps نظیر Versioning ، Automation ، Feature Storing ، CI/CD و Continuous Monitoring . [5]
- ۲- بررسی و انتخاب پلتفرم و ابزارها برای هر قسمت از روند خودکار سازی و بررسی لزوم شخصی سازی آن ها
- ۳- پیاده سازی ابزارهای مورد نیاز به صورت ابری بر روی چندین سرور که خود شامل چندین پردازنده گرافیکی NVIDIA GeForce RTX 3090 می باشد.
- ۴- تست و بهینه سازی ابزارها

مراجع :

- [1] G Symeonidis, E Nerantzis, A Kazakis, “MLOps-Definitions, Tools and Challenges” 2022 IEEE 12th Annual CCWC , 2022
- [2] S Mäkinen, H Skogström, E Laaksonen, “Who needs MLOps: What data scientists seek to accomplish and how can MLOps help?” 2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN), 2021
- [3] DA Tamburri, “Sustainable MLOps: Trends and challenges”, 2020 22nd International (SYNASC), 2020
- [4] <https://mlops-guide.github.io>
- [5] <https://ml-ops.org/>