

دانشکده مهندسی برق

**پایان‌نامه کارشناسی ارشد گرایش سیستم‌های الکترونیک دیجیتال**

طراحی و پیاده سازی یک پلتفرم MLOps به صورت ابری روی GPU

نگارش

**ابوالفضل یاریان**

استاد راهنما

**دکتر متین هاشمی**

... 1402

**برگه تصویب نامه**

تصویب نامه

به نام خدا

دانشگاه صنعتي شريف

دانشکده مهندسی برق

پایان نامه کارشناسي ارشد

این پایان­نامه به عنوان تحقق بخشی از شرایط دریافت درجه کارشناسي­ارشد است.

عنوان: طراحی و پیاده سازی پلتفرم MLOps به صورت ابری روی GPU

نگارش: ابوالفضل یاریان

کميته ممتحنین:

استاد راهنما: دکتر متین هاشمی امضاء.....................................

استاد راهنماي همکار:........................ امضاء.....................................

استاد مشاور:...................................... امضاء.....................................

تاريخ:......................................

برگه اظهارنامه

تقدیم به

این صفحه برای تقدیم پایان‌نامه از سوی دانشجو به افراد یا سازمان‌هایی که برای او مهم هستند در نظر گرفته شده است. اگر پایان‌نامه به کسی تقدیم نمی‌شود این صفحه را پاک کنید.

سپاسگزاری

این صفحه برای سپاسگزاری دانشجو از افراد یا سازمان‌ها در نظر گرفته شده است. اگر از کسی سپاسگزاری نمی‌شود این صفحه را پاک کنید.

# چکیده

در ابتدا هدف از پژوهش و روش پژوهش خود را بیان کنید. در پایان به نتایج و یافته‌های پژوهش اشاره کنید. تعداد کلمات چکیده بین 150 تا 250 کلمه بوده و در یک پاراگراف تنظیم می­شود. از بکارگیری مخفف ها و ذکر مآخذ در چکیده خودداری کنید.

کلیدواژه‌ها: کلیدواژه‌ها را این‌جا وارد کنید. (تعداد کلیدواژه ها 5 تا 7 کلمه)

فهرست مطالب

چکیده ‌ه

فهرست جدول‌ها ‌ح

فهرست تصویرها ‌ط

فهرست نمودارها ‌ي

فصل1 : عنوان فصل اول را اینجا وارد کنید 1

1-1 عنوان فرعی نخست را اینجا وارد کنید 1

1-2 عنوان فرعی دوم را اینجا وارد کنید 1

1-2-1 عنوان فرعی تر را اینجا وارد کنید 1

1-2-1-1 سرعنوان فرعی تر را اینجا وارد کنید 1

فصل2 عنوان فصل دوم را اینجا وارد کنید 2

2-1 عنوان فرعی نخست را اینجا وارد کنید 2

2-2 سرعنوان فرعی دوم را اینجا وارد کنید 2

2-2-1 سرعنوان فرعی تر را اینجا وارد کنید 2

2-2-1-1 سرعنوان فرعی تر را اینجا وارد کنید 2

فصل3 عنوان فصل سوم را اینجا وارد کنید 3

3-1 عنوان فرعی نخست را اینجا وارد کنید 3

3-2 عنوان فرعی دوم را اینجا وارد کنید 3

3-2-1 عنوان فرعی تر را اینجا وارد کنید 3

فصل4 عنوان فصل چهارم را اینجا وارد کنید 4

4-1 عنوان فرعی نخست را اینجا وارد کنید 4

4-2 عنوان فرعی دوم را اینجا وارد کنید 4

فصل5 عنوان فصل 5 را اینجا وارد کنید 6

5-1 عنوان فرعی نخست را اینجا وارد کنید 6

5-2 عنوان فرعی دوم را اینجا وارد کنید 6

منابع یا مراجع 7

پیوست1 8

# فهرست جدول‌ها

تصاویر و جداول بایستی دارای کاپشن اتوماتیک باشند. ((Insert Caption و تمام ارجاع به تصاویر و جداول بایستی به صورت Cross-reference باشند تا بتوان به طور اتوماتیک روزآمد کرد. برای انجام این فرایند در ورد 2007 روی گزینه References کلیک شود و بعد سعی کنید کاپشن جدول یا تصویر بسازید بعد جدول و تصویر را وارد نمایید. انجام این روند، فرایند روزآمدسازی و تهیه فهرست جداول و تصاویر را خیلی ساده می‏کند.

[جدول ‏4‑1- نمونه جدول 4](#_Toc524256067)

# فهرست تصویرها

[شکل ‏4‑1- نمونه تصویر 4](#_Toc524254629)

# فهرست نمودارها

[نمودار ‏4‑1- نمونه نمودار 5](#_Toc524254935)

# فصل اول – مقدمه

در این فصل تعریف مسئله MLOps، اجزا و مزایای آن، چالش های اصلی این حوزه و چهارچوب های کلی پایان نامه توضیح داده خواهد شد.

## تعریف مسئله، اجزا و مزایای آن

فرآیندهای پیاده سازی مدل بخشی پیچیده از چرخه عمر پروژه های یادگیری ماشین می باشد. استفاده از متدولوژی DevOps و ابزارهای یکپارچگی مداوم[[1]](#footnote-1) و ارائه مداوم[[2]](#footnote-2) در توسعه نرم افزار موفقیت هایی داشته است. یادگیری ماشین اغلب بخش کوچکی از سیستم های نرم افزاری است، اما هیچ گاه از این متدولوژی و ابزار CI/CD استفاده نکرده است. سیستم های یادگیری ماشین پیچیدگی و مشکلات منحصر به فردی را به پایپلاین CI/CD اضافه می کند استفاده از این متدولوژی در سیستم های یادگیری ماشین به عنوان MLOps شناخته می شود.

MLOps**، که مخفف عملیات یادگیری ماشین[[3]](#footnote-3) است، یک متدولوژی است که هدف آن بهبود کارآیی، قابلیت اطمینان و مقیاس پذیری جریان کارهای یادگیری ماشین در چرخه عمر آن ها می باشد. با افزایش انتشار مدل های یادگیری ماشین برای تصمیم گیری هایی مبتنی بر داده،** MLOps **به عنوان یک شاخصه حیاتی برای پر کردن شکاف بین توسعه[[4]](#footnote-4)، پیاده سازی[[5]](#footnote-5) و بهبود مداوم[[6]](#footnote-6) سیستم های یادگیری ماشین ظاهر می شود.** MLOps **شامل مجموعه ای از روش ها، ابزارها و رویکردهایی است که همکاری، اتوماسیون و نظارت را در چرخه عمر پروژه های یادگیری ماشین تسهیل می کند. این متدولوژی از عناصری چون** DevOps**، مهندسی داده و مهندسی نرم افزار استفاده می کند تا چالش های منحصربه فرد مرتبط با توسعه و راه اندازی مدل های یادگیری ماشین را برطرف کند.**

**اجزای کلیدی یک پلتفرم** MLOps **به صورت زیر می باشد:**

1. **یکپارچگی و ارائه مداوم (**CI/CD**) : این ابزار به منظور اتوماسیون فرآیندهای یکپارچه سازی تغییرات کدها، تست مدل های یادگیری ماشین و راه اندازی آن های در محیط های تولید استفاده می شود. این امر به معناست که سیستم های یادگیری ماشین به صورت سریع و قابل اعتماد توسعه، اعتبارسنجی و عرضه می شوند.**
2. **کنترل نسخه[[7]](#footnote-7): ابزارهای کنترل نسخه، مانند گیت، نقش حیاتی در این پلتفرم ها به دلیل امکان ردیابی تغییرات کدها و مجموعه داده ها دارند. این ابزار همکاری بین اعضای تیم را تسهیل می کنند، سابقه ای از نسخه های کد، داده را ایجاد می کنند و قابلیت تکرارپذیری را فراهم می کنند.**
3. **هماهنگی زیرساخت[[8]](#footnote-8): این ابزارها، مانند کوبرنتیز، به مقیاس پذیری کارآمد، تخصیص منابع و ظرفیت کانتینریزه کردن مدل های یادگیری ماشین کمک می کنند و راه اندازی قابل اعتماد و مقیاس پذیر را تضمین می نمایند.**
4. **نظارت و مدیریت مدل: نظارت بر عملکرد و رفتار مدل های یادگیری ماشین در محیط های تولیدی جزء جنبه های حیاتی این پلتفرم ها می باشند. این شامل پیگیری معیارهای کلیدی، تشخیص ناهنجاری و امکان مداخله پیشگیرانه برای حفظ دقت مدل است. مدیریت مدل شامل نسخه بندی و راه اندازی محیطی برای روزرسانی های مدل می باشد.**
5. **مدیریت داده: شامل پیش پردازش داده ها، نسخه بندی داده ها و تضمین کیفیت ویژگی های استخراج شده از داده ها می باشد.**

**از مزایای استفاده از پلتفرم های** MLOps **می توان به موارد زیر اشاده کرد:**

1. **افزایش بهره وری: در این پلتفرم ها، وظایف تکراری را اتوماسیون می شود که این امر به دانشمندان داده امکان می دهد تا بر روی وظایف اصلی یادگیری ماشین تمرکز کنند. این کار توسعه و راه اندازی را سریعتر و باعث بهبود بهره وری می شود.**
2. **قابلیت مقیاس پذیری و تکرارپذیری: این امر تضمین می کند که سیستم های یادگیری ماشین قادر به کار با مجموعه داده های بزرگ، تقاضای کاربران بیشتر و نیازهای محاسباتی پیچیده باشند.**
3. **بهبود قابلیت اعتماد مدل: این پلتفرم ها تست، اعتبارسنجی و مکانیزم های نظارت دربرمیگیرد و با نظارت بر عملکرد مدل و رفع مشکلات به موقع باعث جلوگیری از خطاها و انحراف داده ها[[9]](#footnote-9) می شود.**

**محصولات محبوب فعلی برای** MLOps **بیشتر متعلق به ارائه دهندگان ابری مختلف مانند آمازون[[10]](#footnote-10)، مایکروسافت[[11]](#footnote-11) و گوگل[[12]](#footnote-12) هستند. خدمات ارائه دهندگان ابری غالبا راه حل قابل قبولی برای شرکت هایی که روی سیستم های نرم افزاری نظارتی یا نرم افزاری هایی با نگرانی های مربوط به حریم خصوصی کاربر کار می کنند نیستند. آن ها به راه حل هایی نیاز دارند که که قابل اجرا روی ابرها و دستگاه های داخلی باشد. علاوه براین، این خدمات نیز برای کاربران ایرانی دردسترس نمی باشد. از طرف دیگر، راه حل های منبع باز[[13]](#footnote-13) نیز که برای این منظور استفاده می شوند اغلب ناقص و سخت می باشند.**

**هدف پایان نامه، طراحی و پیاده سازی یک پلتفرم** MLOps **برای سیستم های ابری محلی می باشد که نمی توانند به پلتفرم های معروف ارائه دهندگان ابری اتکا کنند. ما یک پلتفرم** MLOps **منبع باز و ساده بر پایه کوبرنتیز که دارای ابزارهای** CI/CD **و نظارت بر سیستم های یادگیری ماشین می باشد ارائه می دهیم. هم چنین ...**

## عنوان فرعی دوم را اینجا وارد کنید

متن را اینجا وارد کنید

### عنوان فرعی تر را اینجا وارد کنید

متن را اینجا وارد کنید

#### سرعنوان فرعی تر را اینجا وارد کنید

# عنوان فصل دوم را اینجا وارد کنید

## عنوان فرعی نخست را اینجا وارد کنید

متن را اینجا وارد کنید

## سرعنوان فرعی دوم را اینجا وارد کنید

متن را اینجا وارد کنید

### سرعنوان فرعی تر را اینجا وارد کنید

متن را اینجا وارد کنید

#### سرعنوان فرعی تر را اینجا وارد کنید

# عنوان فصل سوم را اینجا وارد کنید

## عنوان فرعی نخست را اینجا وارد کنید

متن را اینجا وارد کنید.

## عنوان فرعی دوم را اینجا وارد کنید

متن را اینجا وارد کنید.

### عنوان فرعی تر را اینجا وارد کنید

متن را اینجا وارد کنید.

# عنوان فصل چهارم را اینجا وارد کنید

## عنوان فرعی نخست را اینجا وارد کنید

## عنوان فرعی دوم را اینجا وارد کنید

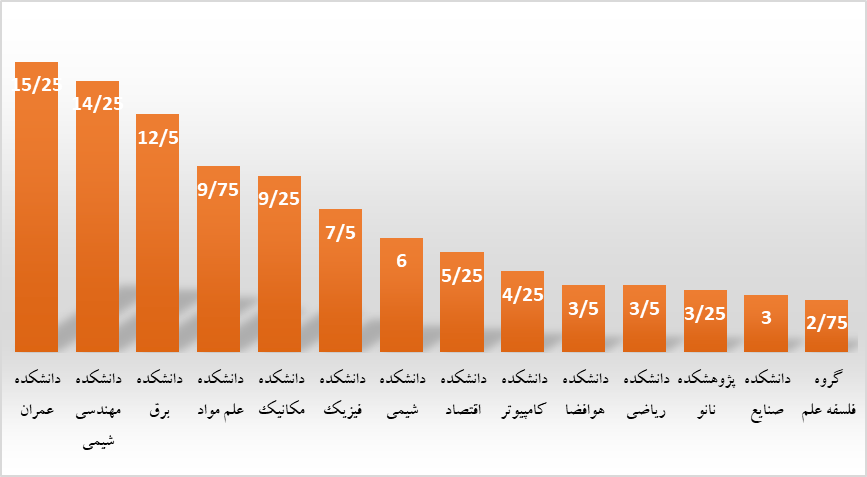
متن را در اینجا وارد کنید (شیوه ارجاع شکل یا جدول در متن جدول ‏4‑ 1) (شکل ‏4‑1)...

جدول ‏4‑1- نمونه جدول

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **رديف** | **ملاک‌ها** | | **ضريب آلفا** | |
| **تحليل پرسشنامه اصلی** | **تحليل پرسشنامه فرعی** |
| 1 | **تمايل به استفاده** | | 7505/0 | 7108/0 |
| 2 | **رضايت‌مندي** | | 7965/0 | 7430/0 |
|  | 2-1 | **مفيدبودن** | 6860/0 | 9216/0 |
| 2-2 | کارآمدي | 8163/0 | 6849/0 |



شکل ‏4‑1- نمونه تصویر

نمودار ‏4‑1- نمونه نمودار

# عنوان فصل 5 را اینجا وارد کنید

## عنوان فرعی نخست را اینجا وارد کنید

متن را اینجا وارد کنید

## عنوان فرعی دوم را اینجا وارد کنید

متن را اینجا وارد کنید

# منابع یا مراجع

(دو شیوه ماخذنویسی وجود دارد به این ترتیب که 1) می­توانید مراجع خود را به ترتیب الفبایی نام خانوادگی نویسنده تنظیم نمایید، در این صورت مراجع فارسی در ابتدا و سپس مراجع لاتین می آید، 2) به ترتیب عددی و براساس حضور مراجع در متن تنظیم نمایید. بدین معنی که اولین مراجعی که در متن استفاده کردید عدد یک و ادامه پیدا می کند. ولی ترجیح با تنظیم الفیایی می باشد.

نام خانوادگی، نام (سال). عنوان و مشخصات کتابشناختی منبع.

نام خانوادگی، نام (سال). عنوان و مشخصات کتابشناختی منبع.)

برای مثال (به ترتیب الفبایی):

* کریمی، مهدی (1387). بررسی تولیدات علمی نویسندگان دانشگاه های ایران همراه با شبکه های تالیف مشترک. فصلنامه علمی و پژوهشی مدیریت. 11 (34): 59-67.
* نیکزاد، مهسا (1389). بررسی تطبیقی شبکه های هم تالیفی در مقالات ایرانی رشته های کتابداری و اطلاع رسانی، روانشناسی، مدیریت و اقتصاد در پایگاه ISI بین سال های 2000 تا 2009. پایان نامه کارشناسی ارشد. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران.
* ولایتی، خالید (1387). بررسی میزان همکاریهای علمی بین ایران و کشورهای همچوار طی سال های 1990 تا 2007. پایان نامه کارشناسی ارشد. دانشگاه تهران.
* Abramo, G. et al (2009). Research collaboration and productivity: is there correlation? High Education, 57:155–171.
* Braun, T. & Glanzel, W (2001), Publication and cooperation patterns of the authors of neuroscience journals, *Scientometrics*, 51 : 499–510
* Chakrabarti, D. & Faloutsos, C. (2006), Graph mining: Laws, generators, and algorithms, ACM Computing Surveys, 38 :1–69.
* Cho, Cheng-Chung, Hu, Ming-Wen, & Liu, Meng-Chun (2010). Improvements in productivity based on co-authorship: a case study of published articles in China. Scientometrics 85: 463-470.
* Durden, G., & Perri, T. J. (1995). Co-authorship and publication efficiency. Atlantic Economic Journal, 23(1):69–76.
* Englebrecht, T. D. et al (2008). An assessment of patterns of co-authorship accountants within premier journals: evidence from 1979-2004. Advances in International Accounting, 24: 172-181.



Abstract

The abstract is the most important section of the thesis because many readers limit most of their reading to abstracts, saving in-depth reading for specific theses. It should give the reader a "preview" of what's to come. The abstract should emphasize new and important aspects of the study or observations. The purpose of the abstract is to allow researchers to decide whether or not to read the whole thesis. The abstract is what researchers read first to decide if the thesis is important, interesting, and it allows them to assess the relevance of a thesis to their own research, without having to read the entire thesis. Thus, it is crucial that the abstract both summarize succinctly the key findings of the thesis and clearly articulate what is novel and important about the work. It should be able to stand alone as a representation of the research without any footnotes. Any information included in the abstract must also be included in the body of the thesis. Although the abstract comes first in the thesis, it is usually easiest to write the abstract last after completing the other sections of the thesis. It is one of the most difficult sections to write. The abstract should generally not contain non-standard acronyms or abbreviations and should not include citations: Briefly, the abstract, introduces topic; mentions techniques used without going into experimental detail; summarizes most important results; is most appropriately written last.

**Keywords (5 to 7 keywords):**

****

**Sharif University of Technology**

**(****Faculty Name)**

Thesis Title

By:

Student Name

Supervisor:

Dr.

Advisor:

Dr.

Month, Year

1. Continuous Integration (CI) [↑](#footnote-ref-1)
2. Continuous Delivery(CD) [↑](#footnote-ref-2)
3. Machine learning and Operation [↑](#footnote-ref-3)
4. Development [↑](#footnote-ref-4)
5. Deployment [↑](#footnote-ref-5)
6. Continuous improvement [↑](#footnote-ref-6)
7. Version Control [↑](#footnote-ref-7)
8. Infrastructure Orchestration [↑](#footnote-ref-8)
9. Data Drift [↑](#footnote-ref-9)
10. AWS SageMaker [↑](#footnote-ref-10)
11. Azure Machine learning [↑](#footnote-ref-11)
12. GCP Vertex AI [↑](#footnote-ref-12)
13. Open-Source [↑](#footnote-ref-13)