In the name of god

Presentation content

1. خوش آمد گویی

|  |
| --- |
| از این که مجدد شما رو می بینیم بسیار خوشحالم و خیلی ممنونم از این که من رو برای این مرحله شورت لیست کردید.  اگر اجازه بدید من دستک تاپم رو برای توضیح درمورد سولشنم شیر کنم |
|  |

1. نشان دادن دو اسلاید که ما دو تا آنالیز انجام دادیم با چه هدفی چرا ؟ و برای اینها از چه زیرساختی استفاده کردیم چرا .. و در آخر توضیح این که ابتدا ما به طور مختصر زیرساخت ها رو توضیح میدیم و بعد میریم سراغ تحلیل eda و بعدش تحلیل های یادگیری ماشین رو توضیح میدیم

|  |
| --- |
| خوب همانطور که می بینید من این تسک رو در دو بخش آنالیزهای توصیفی و پریدیکتویی و یا روش های مبتنی بر یادگیری ماشین انجام دادم  همچنین از دو زیرساخت برای پیاده سازی و اجرای کدها استفاده کردم :  ۱-  ۲-  هدف از داکر بیس اجرای کد توسط شما و دیدن نتایج بوده  در واقع از دو ایمیج تشکیل شده که شامل سه کانتینر هست .. همانطور که در فایل این می بینید یک کانتینر مستر و دو کانتینر وورکر  من برای اسپارک از ایمیج آماده استفاده کردم ولی کانتینر مستر رو طوری کانفیگ کردم که ژوپیتر رو برروی اسپارک اجرا کنه  این کانتینر ها در ریپازیتوری شخصی من به اسم نارسیس در درسترس هست و شما می تونید از طریق ران این کد در ترمینال اون رو دانلود و سپس استفاده کنید.  پس از اجرای کد، یک خروجی برای شما تولید می کند که در واقع ادرس url مربوط به ژوپیتر هست اون رو لانچ می کنید و همانطور که می بینید کدهای مربوط به اینا در دسترس شما هستند.  خوب اجازه بدید بریم در مورد زیرساخت دیگرمون که به طور مستقیم من ازش برای پیاده سازی کدها استفاده کردم صحبت کنیم.  I have set up a five-node cluster using the Hortonwork Data platform which is an open source tool for distributed storage and processing of large, multi-source data.  Actually a have utilized the apache ambari to manage and integrate diffirenet services. Ambari is an openspurce tool which enables system administrators to provision, manage and monitor a Hadoop cluster, and also to integrate Hadoop with the existing enterprise infrastructure.Ambari enables system administrators to provision, manage and monitor a Hadoop.  cluster, and also to integrate Hadoop with the existing enterprise infrastructure.  As you see I have setup a lambda architecture which supports realtime and batch processing task.  For example different services apache ..  which for this task as you requested I have use spark version 2.  ok lets to start with descriptive analysis |
|  |

1. توضیح خلاصه در مورد هورتون وورک – معماری لامبدا – امباری – نشان دادن سرویس ها و گفتن این که از اسپارک استفاده شده
2. توضیح خلاصه در مورد داکر و اسکریپتی که نوشتیم ... طریقه ی ساز و کارش و این که شما می تونید به راحتی ازش استفاده کنید و بعدش گفتن این که خوب تحلیل eda رو شروع کنیم .. اسکریپت رو اجرا می کنیم و بعدش وارد ژوپیتر می شیم
3. توضیحات eda ...

۵-۱ ابتدا فایند اسپارک و ... spark context چیه و چطوری لود شده ؟

۵-۲ به طور خلاصه با نام ذکر توابع عبور می‌کنیم

۵-۳ نحوه ی خواندن داده ها از اچ دی اف اس

۵-۴ پی پراسسیسنگ

۵-۴-۱ رینم کردن ستون ها با استفاده از تابع with column

۵-۴-۲ اکسترنال وریبل ها و تاثیرشون و توضیح توابع udf چی هستن مبتنی بر سطر

۵-۴-۳ ستون تخمین کیفیت – چرا و چطوری ؟

۵-۴-۵ خروجی این تابع و ..

ریزالت ها رو انتهای هر تیکه کد بنویسیم و کپی کنیم

6 distinct توابع توضیح داده شود

۷ سامری

۸ کورلیشن چیه و چی رو نشون میده از چه تابعی استفاده کردم و و بعد میگیم و توضیحات و اینجا انگار خبرایی هست

۹ توزیع ابزار مناسبی است برای این که بدونیم داده ها چطور دیستریبیوت شدن و این شکل ها نشون میده که فلانی و ... بعد از این توابع استفاده کردیم که پلات لی هست

۱۰ ریسمپل کردن چیه چه کمکی می کنه و ... ما این تابع رو با استفاده از اینا نوشتیم و خروجی سه بعدی این شده

۱۱- بررسی تاثیر بارندگی بعد می گیم اجازه بدید که pdf رو نشون بدم که تصاویر متعدد این رو نشون میشده اونجا تاثیر کلی رو نشون میدیم با اون نمودارها

* بعد اینجا میگیم که این نمودار سه بعدی نشون میده که ما می تونیم این رو کلاستر کنیم

۱۲ – تحلیل با زمان رو شروع می کنیم در کل زمان ها و با ریسمپل کردن در ساعت رسم کردیم و دیدیم که تغییر چندانی نداره بنابراین می تونیم بررسی سری زمانی رو در اسکیل ساعت انجام بدیم

تایم سری پلات فانکشن از این تابع ترسیم استفاده می کنه و بعدش اجازه بدید که این تصاویر رو یکجا در پی دف بررسی کنیم ..

۱۳ – بعد میگیم که دیکامپوزیشن چیه و چرا ازش استفاده می کنیم و ما اومدیم دی کامپوز کردیم که ببینم آیا اینا پری دیکتبل هستند ویه سری روش های تست پریدیکشت هست و دوباره نتایج رو ببیتیم

۱۴ – و در اخر هم یک نمودار میله ای برای تغییرات ph و کشیده شده

۶- آنالیز های ماشین لرنینگی

- روش‌های دیگری برای آنومالی دیتکشن در آن سوپروایزد ها آنومالی دیتکشن وجود داره مانند .. و … ولی به دلیل اسپارک بودن تسک من از این استفاده کردم

* ایزالیشون فورست چیه چطوری کار می کنه
* چطوری به اسپارک اضافه میشه و ...
* پارامترهایی که ست کردیم ایناست
* کدش رو نشون میدیم در vscode و میگیم که یک کلاس پیریپراسسینگ داریم و نحوه ی ران کد رو نشون میدیم و بعدش میگیم اینطوری ران می کنیم با پای اسپارک سابمیت
* بعد می گیم برای این که بتونیم خروجی رو ببینیم میریم تو ژو پیتر و بعدش برمیگیردیم به نتایج خروجی هم این شده و توضیح خروجی

۷ – کلاسترینگ

- ابتدا اینکه چرا ک مینز برای این توزیع داده مناسبه ؟ و چرا ما این رو انتخاب کردیم؟

* اولش که توضیح برای لایبری در ام ال لیب
* جواب به این سؤال که أیا مدل های دیگه ای هم می تونستی استفاده کنی ؟
* کد این بوده و مجدد در وی اس کد و ..
* خروجی

۸ – اما در مورد ترنسفر ما

میگم که من در ارائه ی قبلیم به این روش اشاره کردم و برای این آنالیز اومدم کانسپ پروف کردم از کد بیس فلانی استفاده کردم

ترنسفرمرها اینطورین

بعد تنظیمات که کردیم اینطوری بوده و اجراش رو نشون میدیم رو چی پی یو و ... خروجی نهایی رو تو تصاویری که تولید میکنه