

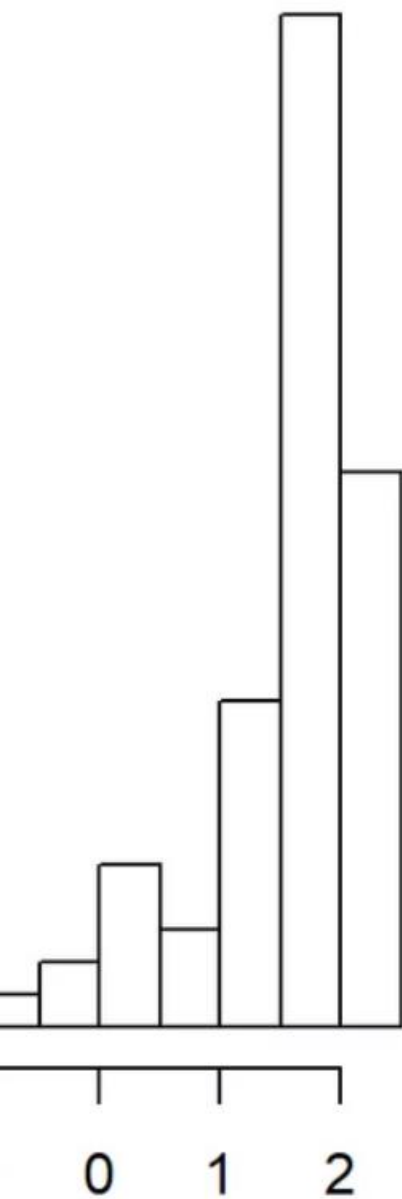
علم داده سطح ۱

الهام حیدری

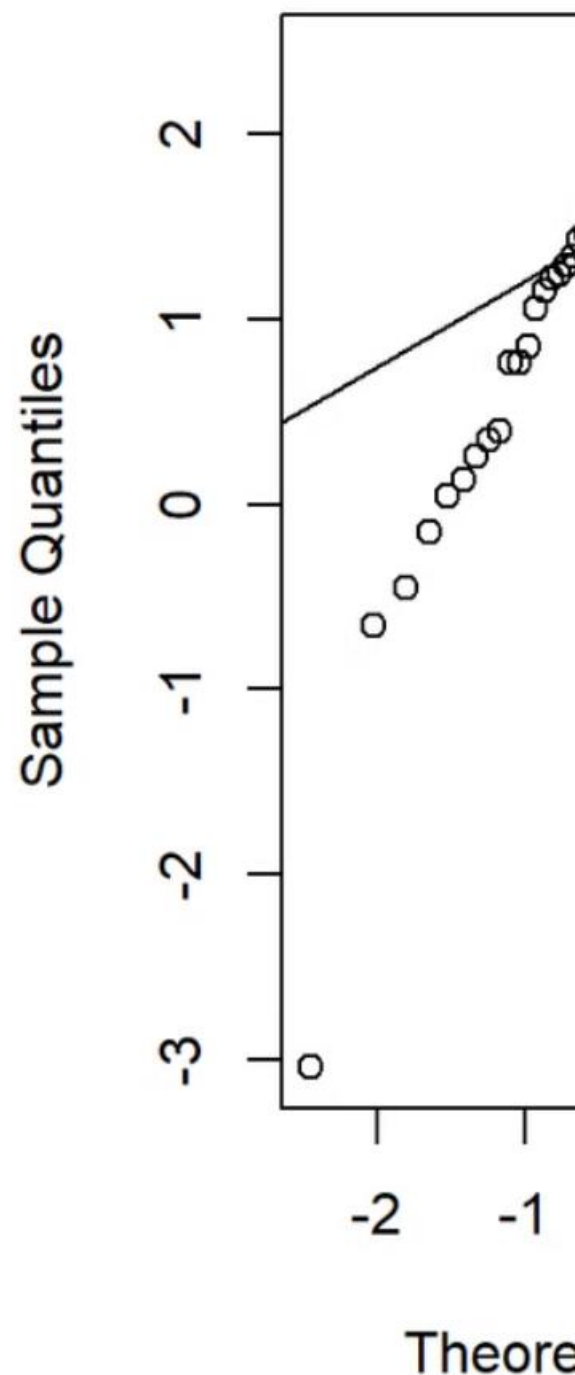
دی ماه ۱۴۰۳



log-normal data



QQ plot for



مقدمه بر علم داده

علم داده یا علم داده‌ها دارای تعاریف مختلفی است اما به طور کلی، این علم شامل تجزیه و تحلیل داده‌های جمع‌آوری شده با هدف استخراج اطلاعات مفید و الگوهای مختلف است. علم داده برای افزایش دانش و ارتقاء تصمیم‌گیری در مواقع مختلف استفاده می‌شود.

E by Elham Heidari

مدرس : الهم حیدری

دانشگاه هاروارد، شغل Data Scientist

جذابترین شغل در قرن ۲۱ معرفی کرد

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>



ن: الهمام حیدری

دیتا ارزشمند ترین منبع دنیا

ارزشمندترین منبع دنیا دیگه نفت نیست بلکه دیتا است

<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>



David Parkins

مدرس : الهمام حیدری

MODERN DATA SCIENTIST

At the moment, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from mathematics, statistics, computer science, communication and business. Finding people who understand who a data scientist is, is equally hard. So here is a cheat sheet on who the modern data scientist really is.

PRO & DA

- ☆ Cor
- ☆ Sci
- ☆ Sta
- ☆ Dal
- ☆ Rel
- ☆ Pai
- ☆ pro
- ☆ Ma
- ☆ Ha
- ☆ Cu
- ☆ Exp

COM & VI

- ☆ Ab
- ☆ ma
- ☆ Ste
- ☆ Tra
- ☆ der
- ☆ Vis
- ☆ R p
- ☆ Kn
- ☆ toc



تعریف علم داده

تحلیل داده

علم داده شامل تحلیل داده‌های ساخته شده با استفاده از روش‌های آماری و الگوریتم‌های تحلیلی می‌باشد که هدف اصلی آن استخراج اطلاعات و الگوهای کاربردی است.

مهارت‌های تحلیلی

یک دانشجوی علم داده نیازمند داشتن مهارت‌های تحلیل داده و تفسیر اعداد، تجربه برنامه‌نویسی و آشنایی با الگوریتم‌ها و سیستم‌های مبتنی بر داده است.

وظایف متعدد

مهم‌ترین وظایف یک دانشمند داده شامل تحلیل، پیش‌بینی و رفع مشکلات است که نیازمند همه‌جانبه بررسی داده‌ها و مطالعه‌ی آنها می‌باشد.

کاربردهای علم داده

1

تصمیم‌گیری بهتر

اطلاعات به دست آمده از علم داده می‌تواند به شرکت‌ها در اتخاذ تصمیمات بهتر و دقیق‌تر کمک کند.

2

بهبود تجربه کاربران

استفاده از داده‌ها در بهتر کردن و بهبود جوامع، محصولات یا خدمات به مشتریان تاثیر مثبت خواهد داشت.

3

پیش‌بینی دقیق

با استفاده از الگوریتم‌ها و مدل‌های علم داده می‌توان پیش‌بینی‌های دقیقی ارائه داد.



دلیل رشد علم داده؟



- تولید حجم انبوهی از داده
- افزایش توان سخت افزاری برای ذخیره سازی این حجم از داده ها با قیمت پایین
- پیشرفت روش های محاسباتی و الگوریتم ها برای تجزیه و تحلیل داده های با حجم بالا



5 MB Hard Disk Drive - 1956

© Go Physics



1 TB Micro SD Card - 2020



فرآیند تحلیل داده

تفسیر

مرحله‌ی آخر فرآیند تحلیل داده شامل تفسیر و تحلیل نتایج به دست آمده به صورت کامل و جامع می‌باشد.

مدل سازی

در این مرحله، مدل‌هایی برای تفسیر داده‌ها و ارائه گزارشات تحلیلی تر ساخته می‌شود.

تجزیه و تحلیل

مرحله‌ی اول فرآیند تحلیل داده شامل تجزیه و تحلیل داده‌های موجود است.

مراحل استخراج داده

1

تجميع داده‌ها

2

پیش پردازش داده‌ها

3

تحليل داده‌ها

4

ارائه گزارش نتایج

مهارت های ورود به علم داده



ریاضیات و آمار
برنامه نویسی پایتون یا R
پایگاه داده SQL
نمایش داده
یادگیری ماشین
یادگیری عمیق
دانش تخصصی حوزه
مهارت های ارتباطی و ارائه
استدلال کردن

تحلیل و تفسیر داده ها

1

تجزیه و تحلیل داده‌ها

نقشه‌ی برداری و بررسی داده‌ها بهره‌مند
از روش‌های مدرن علم داده

2

ارزیابی و تحلیل

مرحله‌ی مهم تحلیل اطلاعات جمع‌آوری
شده برای شناخت بهتر الگوها و تبیین و
فهم دقیق داده‌ها

3

تفسیر و نتیجه‌گیری

با بررسی و تفسیر نتایج، تصمیمات بهتر و
دقیق‌تری می‌توان اتخاذ کرد

الگوریتم ها و مدل های علم داده

درخت تصمیم

ماشین بردار پشتیبان

شبکه های عصبی

الگوریتم یادگیری ماشین

الگوریتم کاهش بعد

خوشه بندی کی میانگین

- increased efficiency
- funding
- collaboration/networking

- reuse of funded research
- greater funding impact
- greater ROI

- increased understanding and expertise
- promoting engagement in science & research



Organisations/
NGOs



National
Governments

مزایای استفاده از علم داده

ارتقاء تصمیم‌گیری

معمولاً علم داده به سازمان‌ها کمک می‌کند تا تصمیم‌گیری‌های بهتری انجام دهند؛ گرچه ممکن است نتایج آن همیشه صحیح نباشد، اما تصمیمات مبتنی بر علم داده معمولاً بهتر از تصمیمات مبتنی بر گزاف است.

بازاریابی هدفمند

برنامه‌ریزی موثر و بهینه‌سازی رویکردهای بازاریابی با استفاده از داده‌ها کمک می‌کند که به مشتریان مناسب‌تری دست یابیم و ارزش‌های بیشتری برای شرکت خلق کنیم.

کاهش هزینه‌ها

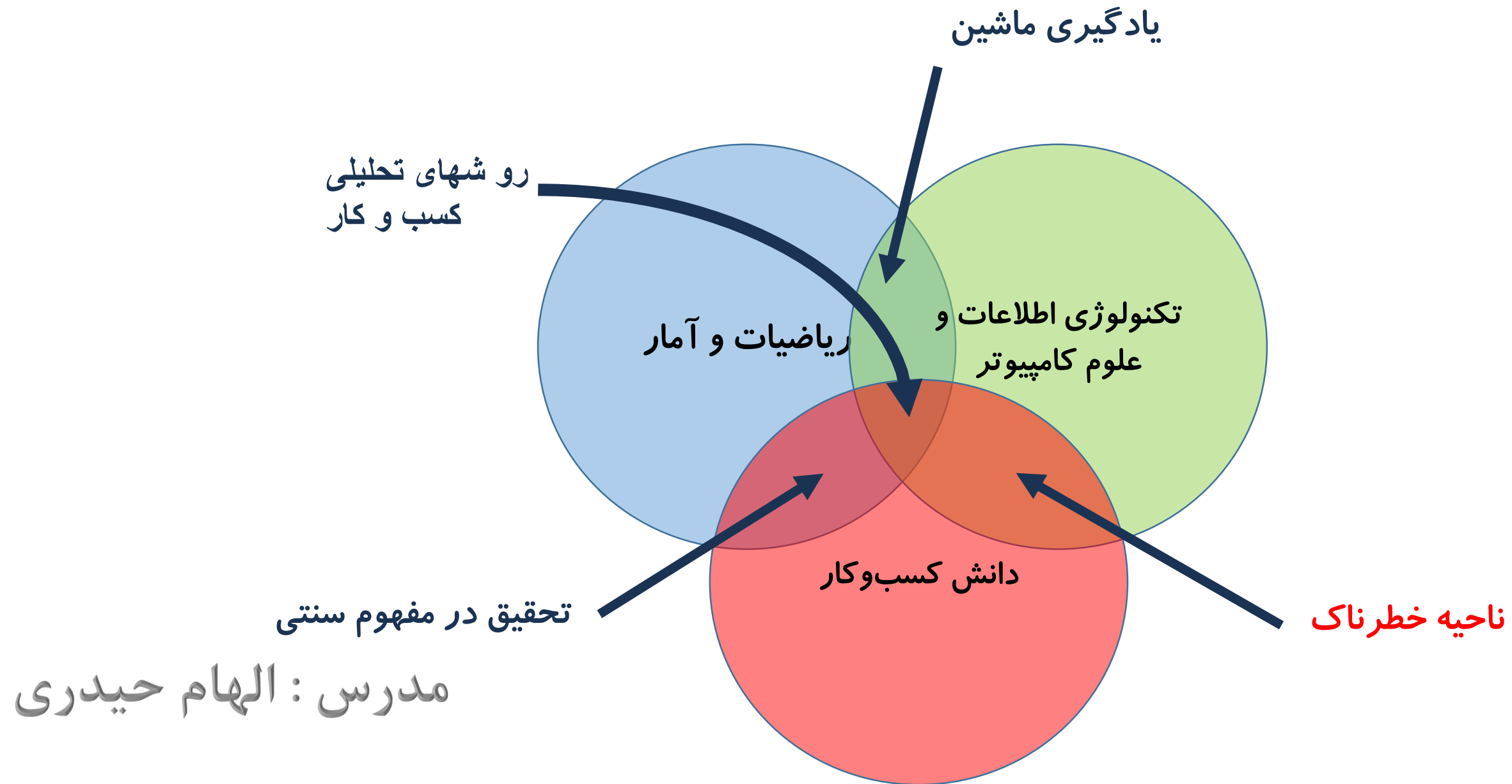
استفاده از داده‌ها و الگوریتم‌های علم داده به شرکت‌ها کمک می‌کند هزینه‌ها را کاهش دهند و یا منابع خود را بهینه‌سازی کنند.

مسیر شغلی جذاب

- تقاضا برای علم داده
- رشد 36 درصدی در آمریکا تا سال 2031
- سطح در آمدی بالا
- نیاز به یادگیری و رشد مداوم
- تغییرات بهتر در دنیا



حوزه های دانشی در علم داده



چالش ها و مشکلات علم داده



1

حجم بزرگ داده‌ها

مدیریت و تجزیه و تحلیل داده‌های بزرگ می‌تواند چالش بزرگی برای سازمان‌ها و تحلیل‌گران باشد.

2

پایداری داده‌ها

به دست آوردن داده‌های پایدار و قابل اعتماد می‌تواند مشکلی برای استفاده‌ی عمومی باشد.

3

محافظت از حریم خصوصی

حریم خصوصی و امنیت داده‌ها از چالش‌های اصلی در علم داده محسوب می‌شود.

مشاغل علم داده

- دانشمند داده Data Scientist
- مهندس داده Data Engineer
- تحلیلگر داده Data Analyst
- کارشناس هوش تجاری Business Intelligence Developer

- مهندس یادگیری ماشین Machine Learning Engineer
- توسعه دهنده ی هوش مصنوعی Artificial Intelligence Developer
- مدیر تیم داده Data Team Lead

مدرس : الهمام حیدری

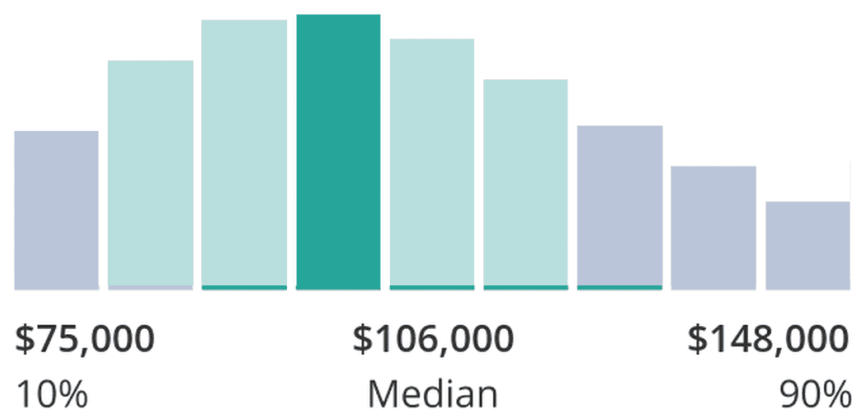
درآمدی دیتا ساینس در جهان

DATA SCIENTIST

Average Salary

\$106,000 yearly

\$51.01 hourly



Updated August 18, 2021

Junior Level

25th Percentile

\$88,000

Yearly Salary

\$42.31

Hourly Salary

Mid Level

50th Percentile

\$106,104

Yearly Salary

\$51.01

Hourly Salary

Senior Level

75th Percentile

\$126,000

Yearly Salary

\$60.58

Hourly Salary

وضعیت علم داده در ایران

برنامه نویس هوش مصنوعی و روباتیک

گروه پیروی

تهران، پاسداران | ۲۵ - ۲۰ میلیون تومان

۲۸ روز پیش



کارفرمای پاسخگو

کارشناس هوش مصنوعی

بهساتران ملت

تهران، پاسداران

۲۲ روز پیش



مدرس علم داده

دفتر بنیاد مدیریت کیفیت اروپا

تهران، شهرک غرب | ۲۵ - ۲۰ میلیون تومان

۴۳ روز پیش



کارشناس هوش مصنوعی

ویرا ۱۱۰

مشهد | ۸ - ۶ میلیون تومان

امکان جذب کارآموز

۴۸ روز پیش



فوری

کارشناس تحلیل داده (Data Analyst)

بیمه پاسارگاد (دفتر مرکزی)

تهران، فاطمی

۳۴ روز پیش



فوری

کارشناس مهندس داده (Data Engineer)

ارتباطات سیار ایران - همراه اول

تهران، ونک

۵۸ روز پیش



۲۲ روز پیش



Junior Data Analyst

استپ فود

تهران، میرداماد

۲۱ روز پیش



کارفرمای پاسخگو

در حال بررسی رزومه ها

کارشناس علم داده

داتیس آرین قشم - داتین

تهران، جردن

۲۸ روز پیش



فوری

کارفرمای پاسخگو

در حال بررسی رزومه ها

کارشناس هوش تجاری

نویان نگین پارسیان

تهران، آیت الله کاشانی | ۲۰ - ۱۵ میلیون تومان

۲۹ روز پیش



کارفرمای پاسخگو

machine learning engineer (مهندس یادگیری ماشین)

شرکت هوش مصنوعی گاتا

تهران، هفت تیر

۳۶ روز پیش



مدرس : الهمام حیدری

سر فصل ها



مقدمه ای بر علم داده



مراحل اجرای یک پروژه داده کاوی

نصب آناکوندا، کار اولیه با پایتون



ساختارهای داده، عبارات شرطی و حلقه ها در پایتون



ماژول ها و کتابخانه، توابع و کلاس در پایتون

آشنایی با کتابخانه Numpy در پایتون،

آشنایی با کتابخانه Pandas در پایتون

سر فصل ها



آشنایی با کتابخانه Matplotlib در پایتون



روش‌های آمار توصیفی و استنباطی برای شناخت داده - موردکاوی اول



نحوه مواجهه با داده‌های گمشده و داده‌های پرت



EDA

مقدمه‌ای بر یادگیری ماشین
الگوریتم رگرسیون خطی

الگوریتم kNN

مراحل مدل‌سازی در یادگیری ماشین - موردکاوی سوم

سر فصل ها



رویکردهای قاعده‌سازی در رگرسیون خطی -



درخت تصمیم و جنگل تصادفی در یادگیری ماشین - مورد کاوی
ششم

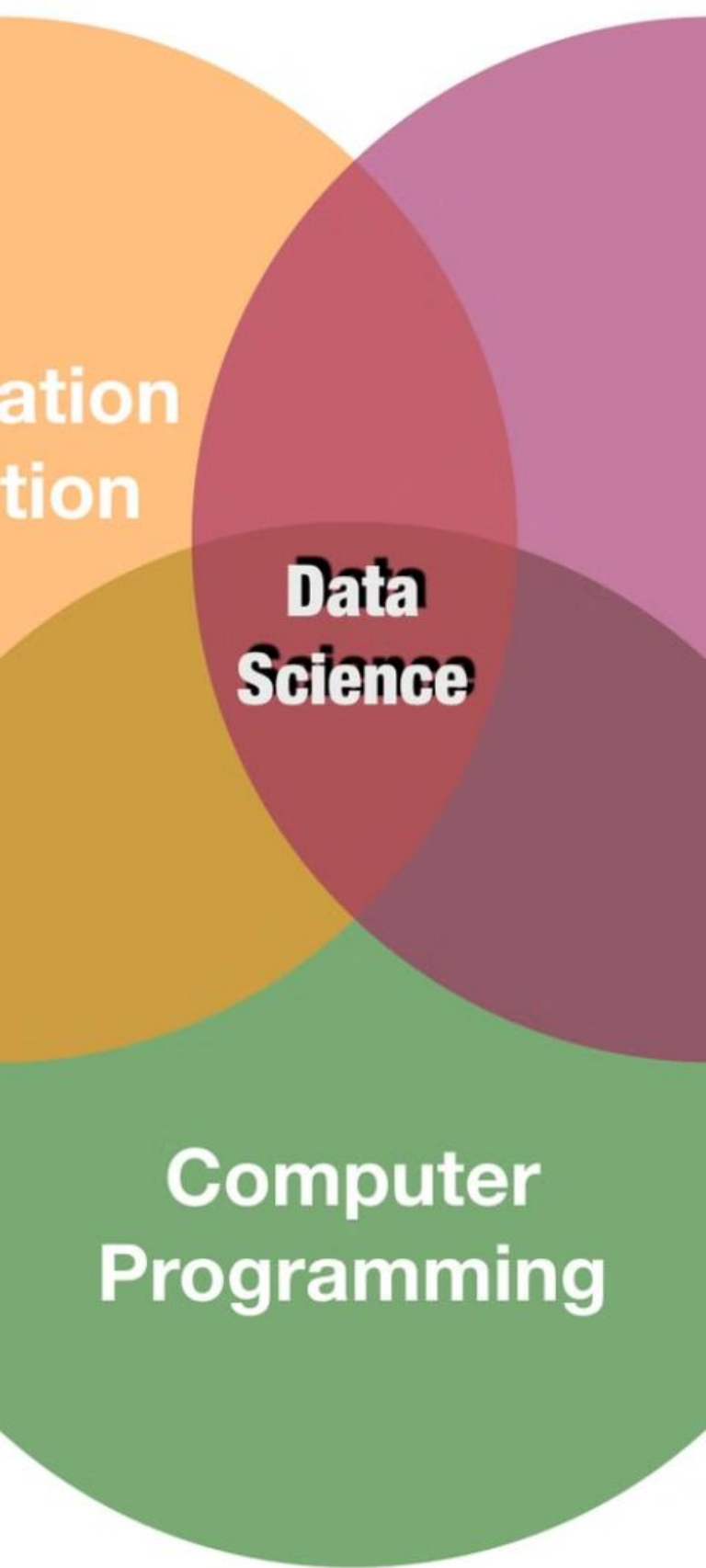


رگرسیون لجستیک



الگوریتم‌های دسته‌بندی در پایتون

الگوریتم‌های نظارت‌نشده در پایتون - مورد کاوی هشتم



نتیجه گیری و پیشنهادات

1

تقویت دانش علم داده

پیشنهادهای ما شامل تقویت دانش مرتبط با فنون و مفاهیم علم داده و استفاده از پایگاه‌های داده‌ها برای مشترکین می‌باشد.

2

توسعه فناوری

ما توصیه می‌کنیم در توسعه فناوری داده‌ها، از جمله سیستم‌های پایگاه داده، استفاده کنید تا به داده‌های خام دسترسی پیدا کنید.

3

انتقال اطلاعات صحیح

توصیه می‌کنیم که اطلاعات به دست آمده از عملیات مختلف تجارتي و پژوهشی را به دقت و دقیق منتقل کنید؛ زیرا این اطلاعات می‌توانند تأثیرات مهمی داشته باشند.

موفقیت

یادگیری و مهارت

انجام پروژه

- github
- linkedin
- kaggle



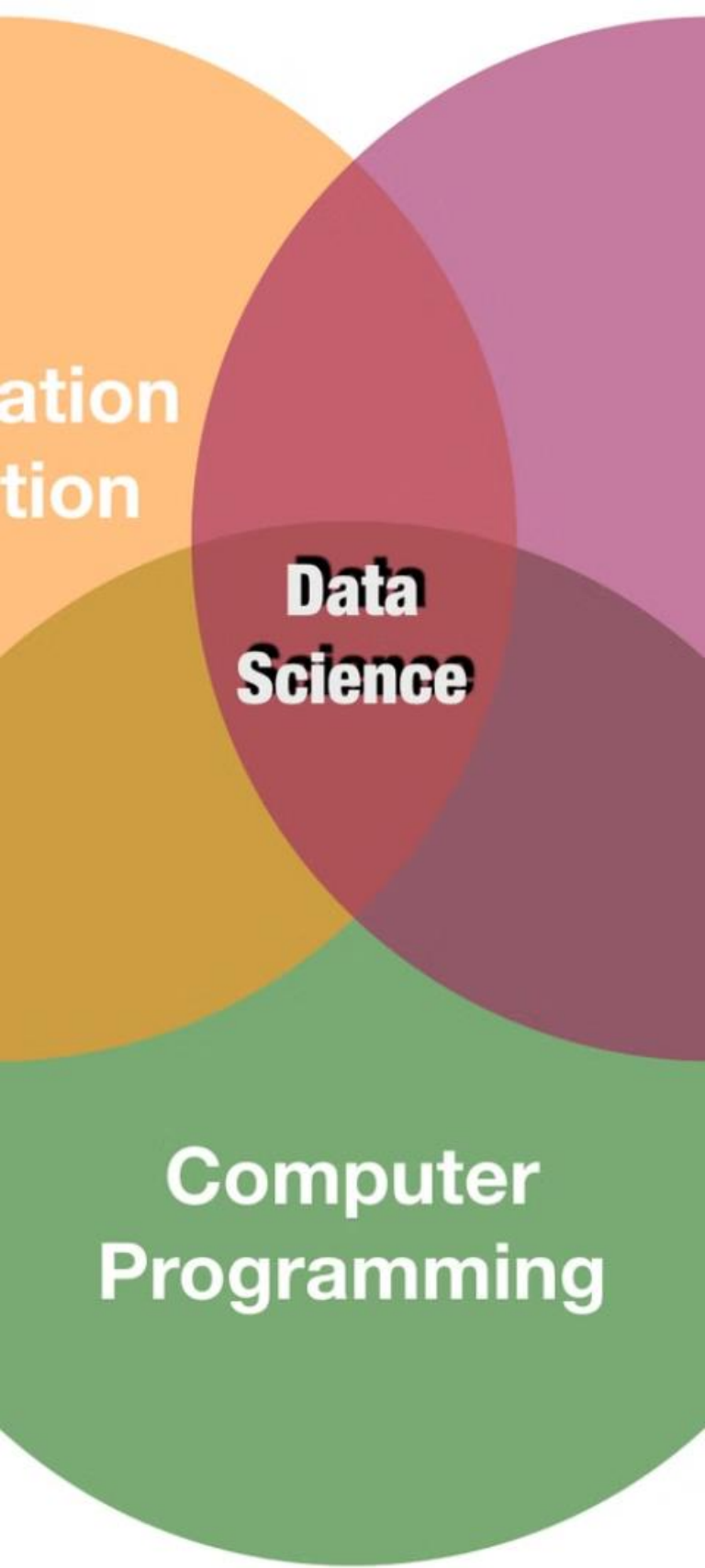
networking

<https://www.linkedin.com/in/heidari-ai>

<https://github.com/elihe90>

<https://www.kaggle.com/eli2022>

مدرس : الهمام حيدري



Thank you

QUESTIONS?