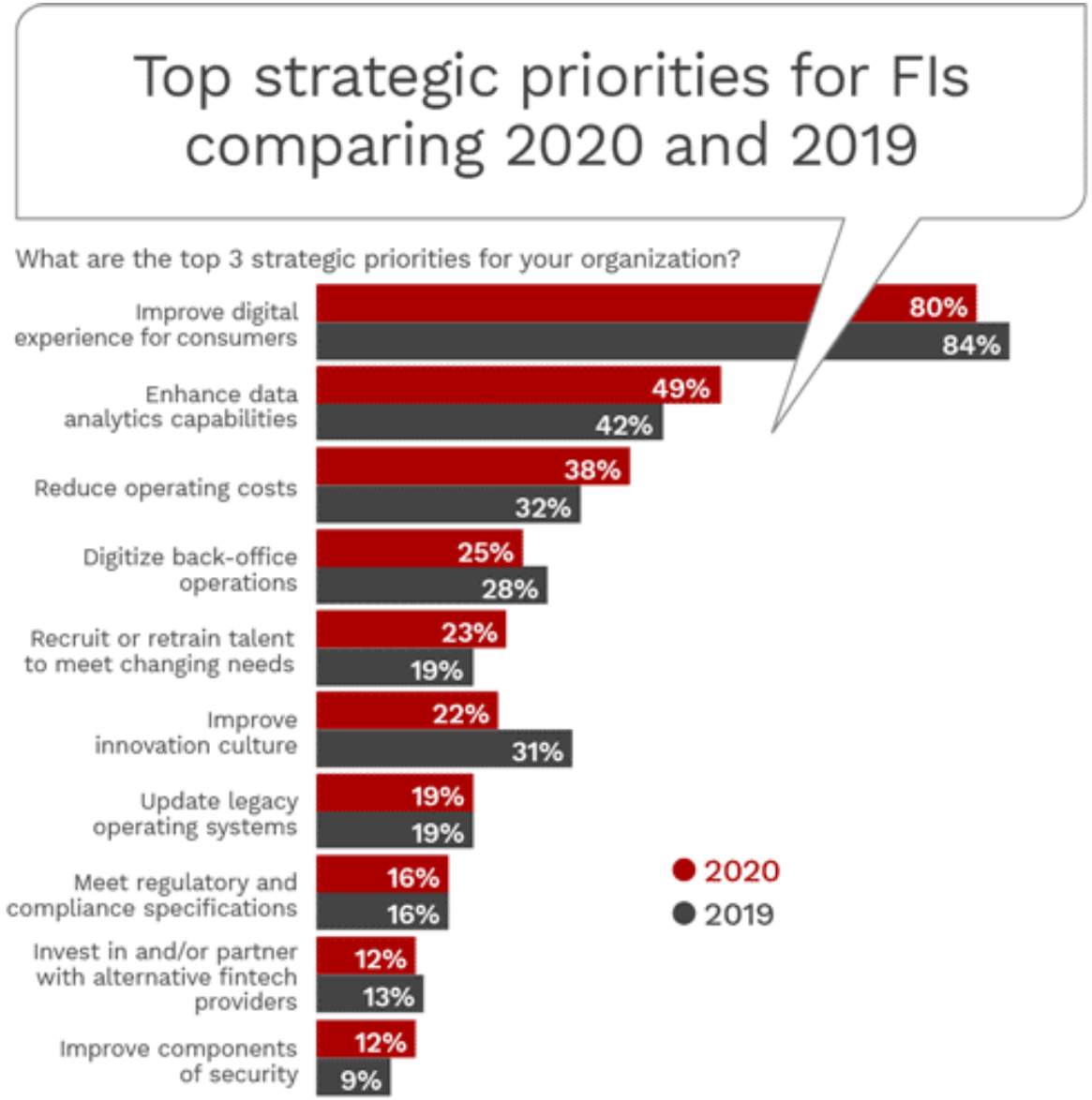
بسم الله الرحمن الرحيم

محمدرضا محتاط

متخصص علم‌داده و مشاور داده‌های عظیم

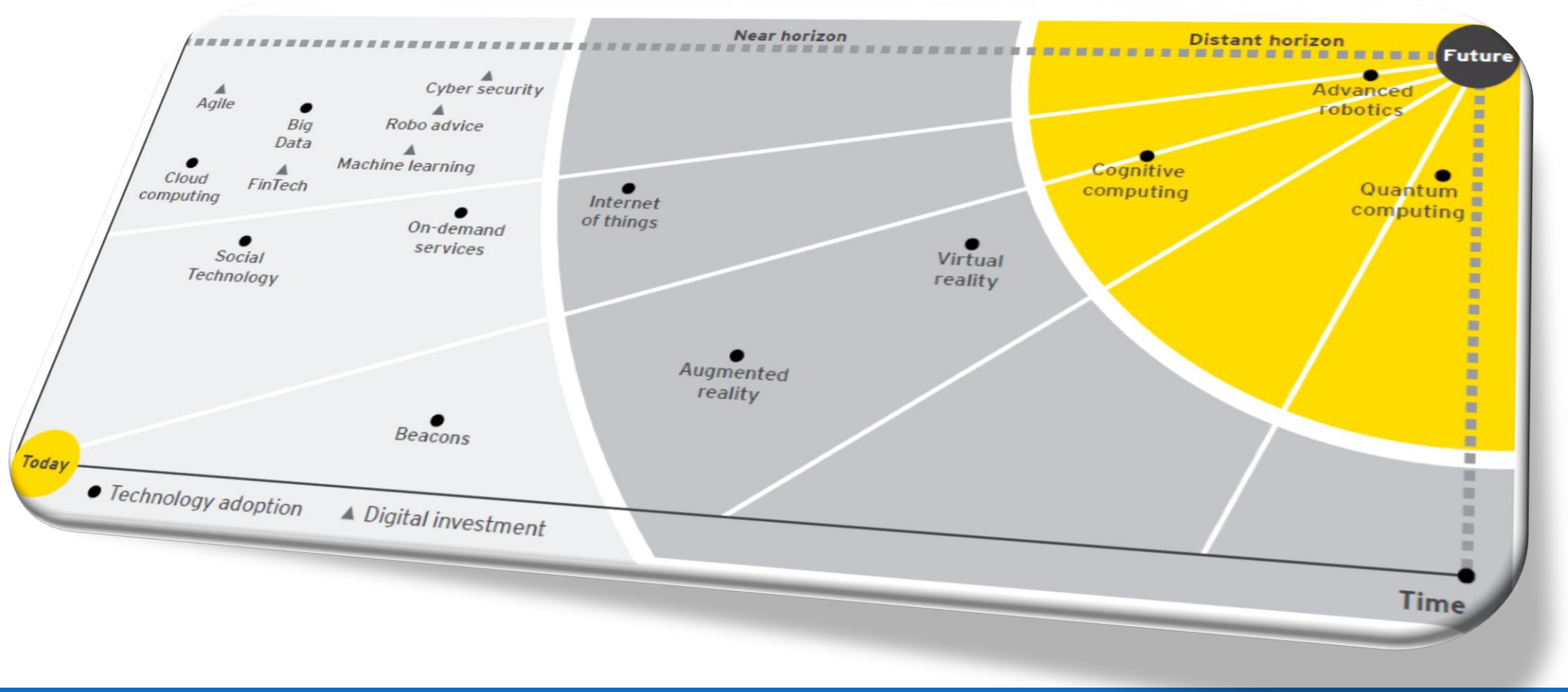
- ❖ عضو کمیته راهبری و اساتید دوره جامع علم‌داده دانشگاه تهران
- ❖ مشاور ارشد پروژه تدوین نظام تحلیل داده‌های بانک ملت
- ❖ پیمانکار علوم داده مرکز ملی فضای مجازی
- ❖ عضو حقوقی کمیسیون علوم داده و هوش مصنوعی سازمان نظام صنفی رایانه‌ای
- ❖ عضو حقیقی کمیسیون اینترنت اشیا و داده‌های عظیم سازمان نظام صنفی رایانه‌ای
- ❖ مشاور اسبق داده‌های عظیم شرکت‌های صنایع یاس ارغوانی، دیتاک و...
- ❖ رتبه برتر مسابقات ملی داده‌کاوی دانشگاه شریف (فناورد)
- ❖ دانشجوی دکتری مدیریت فناوری اطلاعات دانشگاه علامه طباطبایی
- ❖ کارشناسی ارشد مدیریت فناوری اطلاعات دانشگاه تهران
- ❖ مدیریت پیشرفته کسب‌وکار(MBA) گرایش استراژی دانشگاه تهران



SOURCE: Digital Banking Report Research © January 2020 The Financial Brand

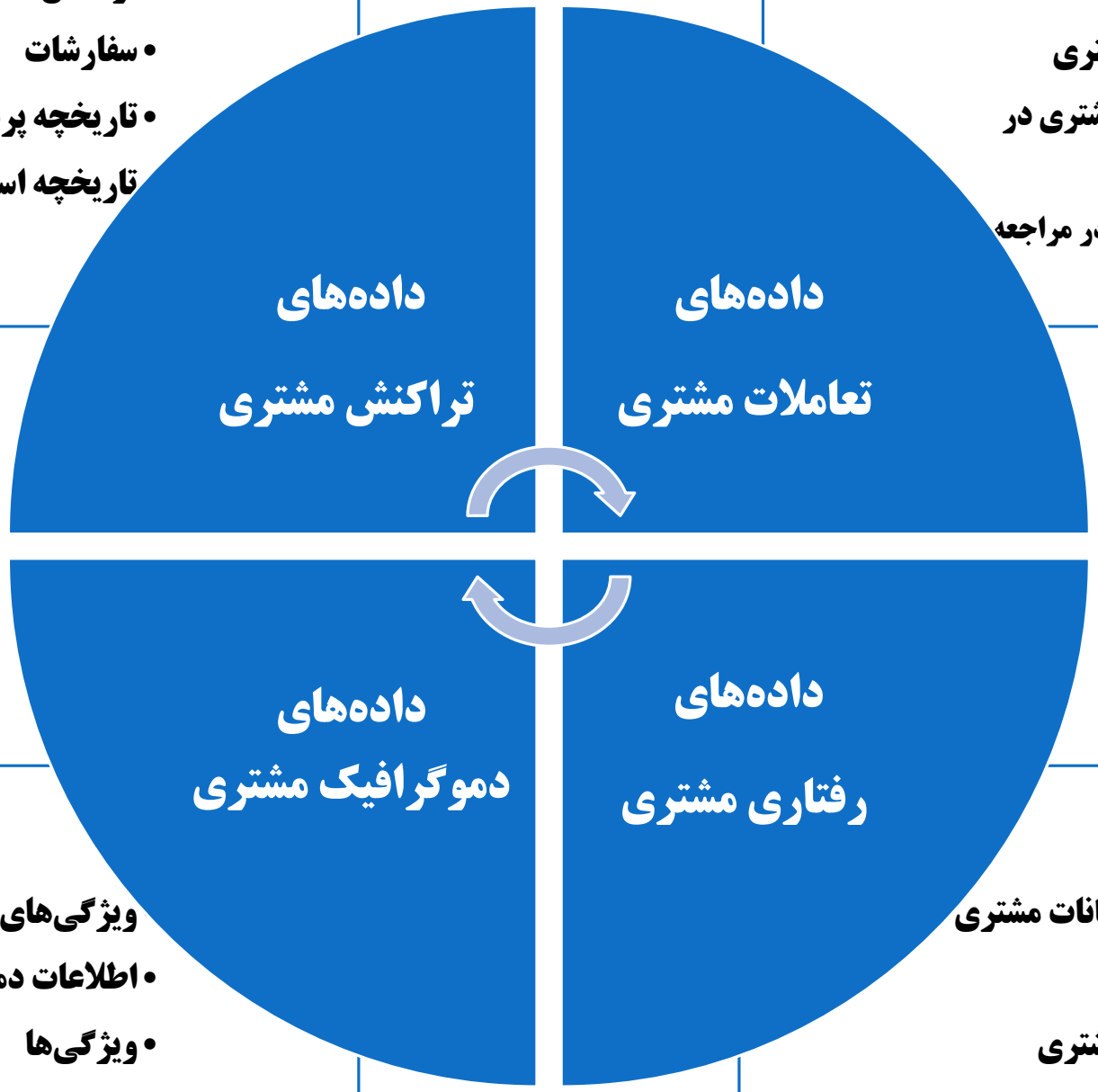
برترین اولویت‌های استراتژیک بانک‌ها در ۲۰۲۰

فناوری‌های تحول آفرین در صنعت بانکداری



کاربردهای علم‌داده در صنعت بانکداری

تحلیل جامع مشتریان



- ایمیل / چت
- مرکز ارتباط با مشتری
- جریان کلیک‌های مشتری در وب
- دیالوگ‌های مشتری در مراجعه حضوری

- تراکنش‌ها
- سفارشات
- تاریخچه پرداخت‌ها

تاریخچه استفاده از خدمات

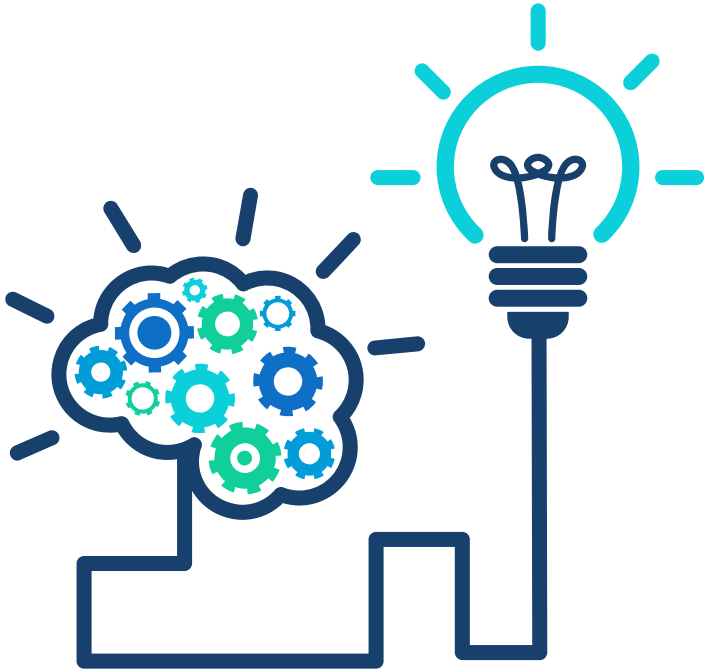
- احساسات و هیجانات مشتری
- تمایلات مشتری
- نیازمندی‌های مشتری
- خواسته‌های مشتری

ویژگی‌های شخصیتی

- اطلاعات دموگرافیک
- ویژگی‌ها

تحلیل مشتریان

- بخش‌بندی مشتریان
- رتبه‌بندی مشتریان
- طبقه‌بندی مشتریان(طلایی، عادی و...)
- ارزیابی اعتبار مشتریان بانک
- محاسبه طول عمر مشتری
- تحلیل میزان وفاداری مشتری
- پیش‌بینی ترک مشتری



خوشه‌بندی مشتریان بانکی

Attached: Spark | File | View: Code | Permissions | Run All | Clear | Publish | Comments | Revision history

Cmd 1

```python
#Step 1: Create the SparkSession Object
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName('K_means').getOrCreate()
```

Command took 0.03 seconds -- by mohtat@ut.ac.ir at 2/10/2019, 7:34:24 PM on Spark

Cmd 2

```python
# File location and type
file_location = "/FileStore/tables/Ghavamin.csv"
file_type = "csv"

# CSV options
infer_schema = "false"
first_row_is_header = "True"
delimiter = ","

# The applied options are for CSV files. For other file types, these will be ignored.
df = spark.read.format(file_type) \
  .option("inferSchema", infer_schema) \
  .option("header", first_row_is_header) \
  .option("sep", delimiter) \
  .load(file_location)

display(df)
```

(2) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [PK: string, GENDER: string ... 8 more fields]

PK	GENDER	AGE	CUSTOMER_TYPE	ACC_IND_DESC	ACC_TYPE_DESC	ACC_OPEN_DATE	A
1.000000	0.000000	33.000000	1.000000	1.000000	0.000000	11.000000	1
2.000000	0.000000	33.000000	1.000000	1.000000	0.000000	11.000000	1
3.000000	1.000000	19.000000	1.000000	1.000000	2.000000	39.000000	0
4.000000	0.000000	64.000000	1.000000	1.000000	2.000000	75.000000	1
5.000000	1.000000	34.000000	1.000000	1.000000	2.000000	32.000000	0
6.000000	0.000000	54.000000	1.000000	1.000000	2.000000	12.000000	0
7.000000	0.000000	54.000000	1.000000	1.000000	2.000000	12.000000	0
8.000000	0.000000	54.000000	1.000000	1.000000	2.000000	4.000000	0

Command took 3.45 seconds -- by mohtat@ut.ac.ir at 2/10/2019, 7:34:25 PM on Spark

Cmd 10

```python
#Step 5: Build K-Means Clustering Model
from pyspark.ml.clustering import KMeans
```

Command took 0.02 seconds -- by mohtat@ut.ac.ir at 2/10/2019, 7:38:12 PM on Spark

Cmd 11

```python
errors=[]
for k in range(2,10):
  kmeans = KMeans(featuresCol='features',k=k)
  model = kmeans.fit(final_data)
  intra_distance = model.computeCost(final_data)
  errors.append(intra_distance)
print(errors)
```

(65) Spark Jobs

```
[4.0579867671993156e+19, 1.1929042247158934e+19, 6.09595915030981e+18, 3.3130475543162143e+18, 1.838640001408791e+18, 9.322627706111657e+17, 8.120845943866721e+17, 5.589248480942568e+17]
```

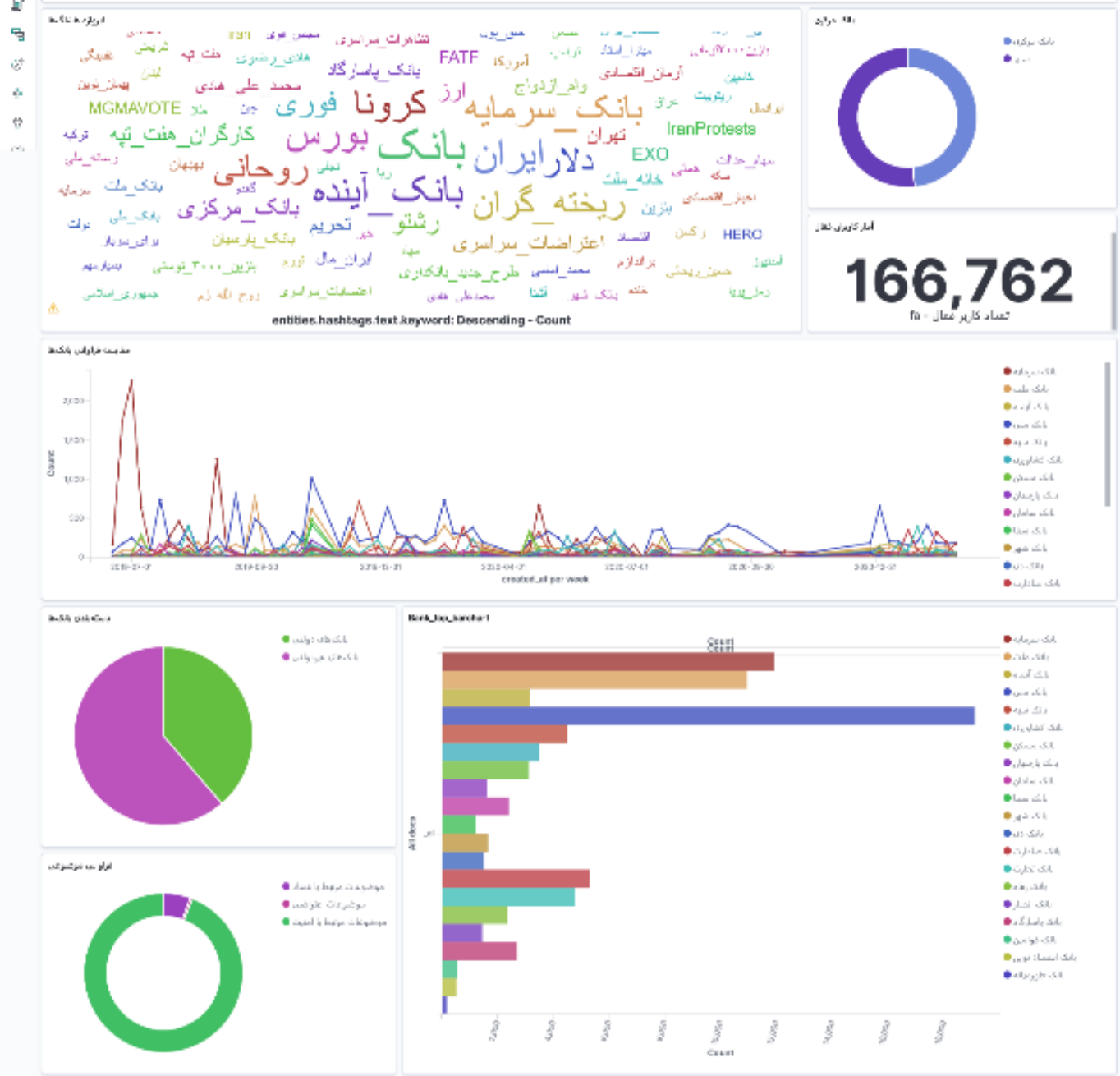
Command took 16.17 seconds -- by mohtat@ut.ac.ir at 2/10/2019, 7:38:15 PM on Spark

Cmd 12

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
cluster_number = range(2,10)
plt.xlabel('Number of Clusters (K)')
plt.ylabel('SSE')
plt.scatter(cluster_number,errors)
display(plt.show())
```

تحلیل مشتریان

- تحلیل احساسات نظرات مشتریان برروی خدمات، کانال‌های پرداخت، شعب و...
- تحلیل سبد و نیازمندی‌های مشتریان
- ارائه محصولات شخصی‌سازی شده (سیستم‌های توصیه‌گر)
- توسعه محصولات جدید بنا به نیازمندی‌های مشتریان
- تحلیل باشگاه مشتریان

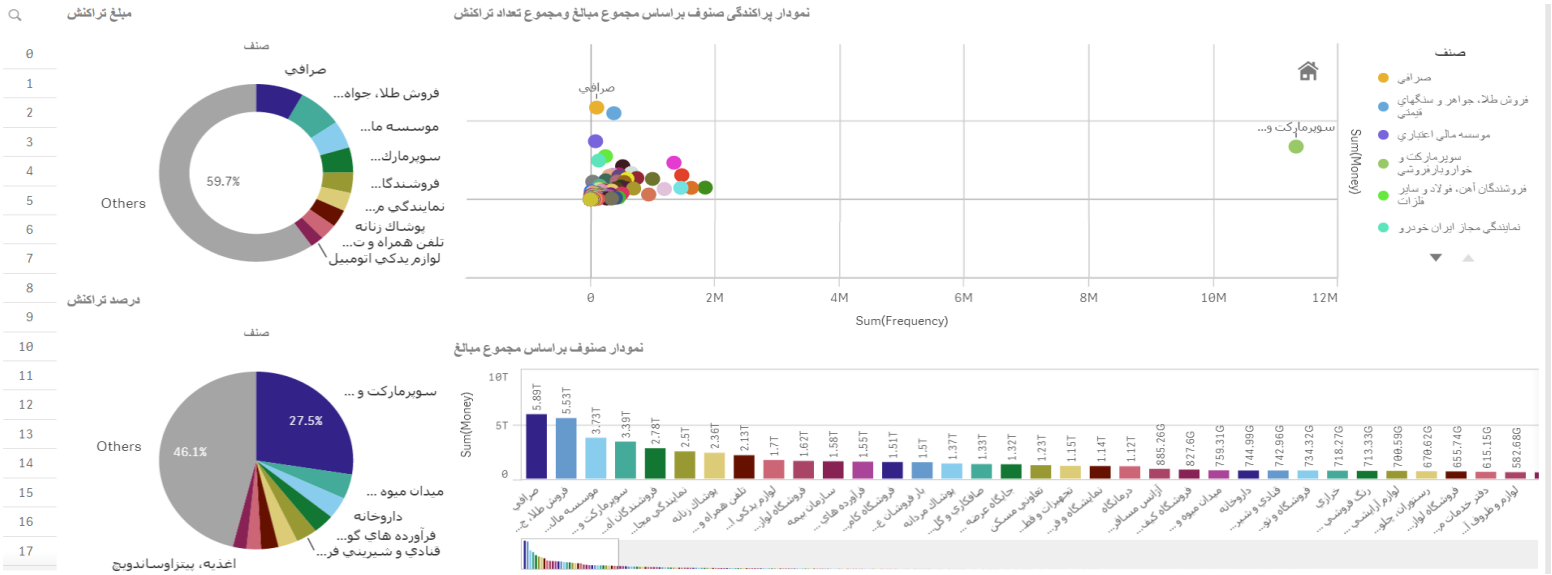


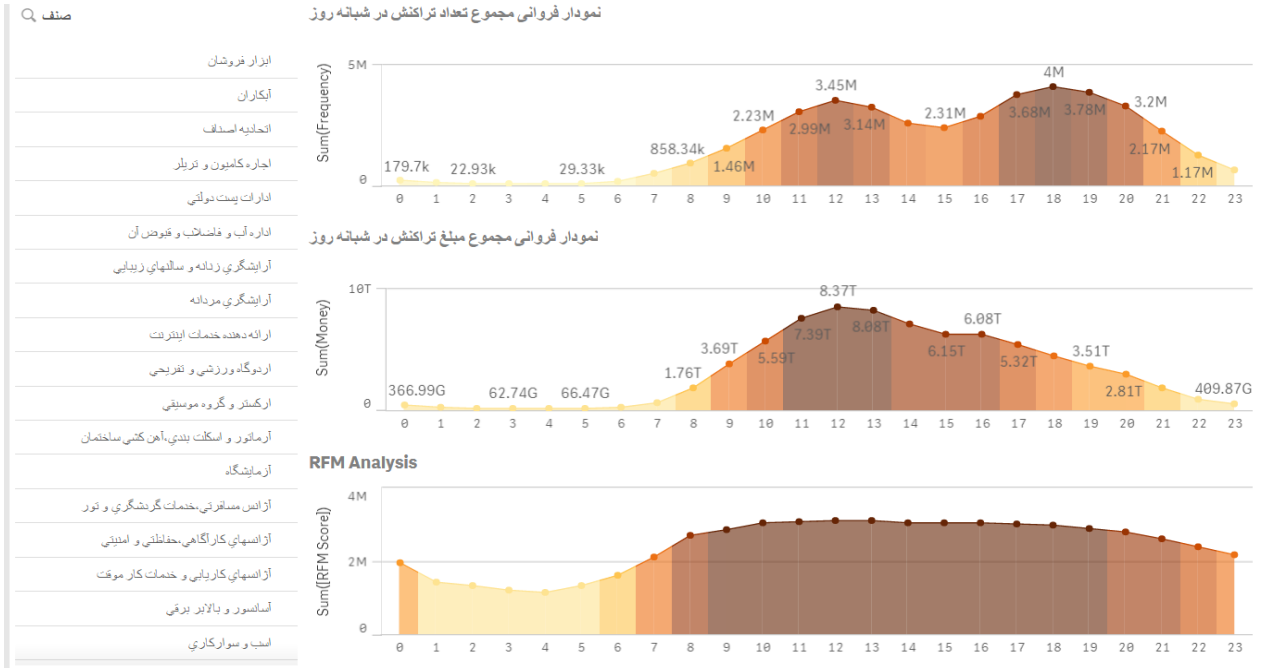
داشبورد تحلیل برند بانک‌ها در شبکه‌های اجتماعی

تحلیل کانال‌های ارتباطی بانک با مشتریان

- پیش‌بینی روزانه مقدار موجودی مورد نیاز در دستگاه‌های ATM
- مکان‌یابی دستگاه‌های خودپرداز جدید و بازطراحی شبکه‌های ATM
- ارزیابی بهره‌وری و تحلیل عملکرد دستگاه‌های خودپرداز
- ارزیابی و تحلیل تراکنش‌ها (تراکنش‌های ناموفق و...)
- برنامه‌ریزی نگهداری و تعمیرات دستگاه‌های خودپرداز

رتبه‌بندی صنوف تهران





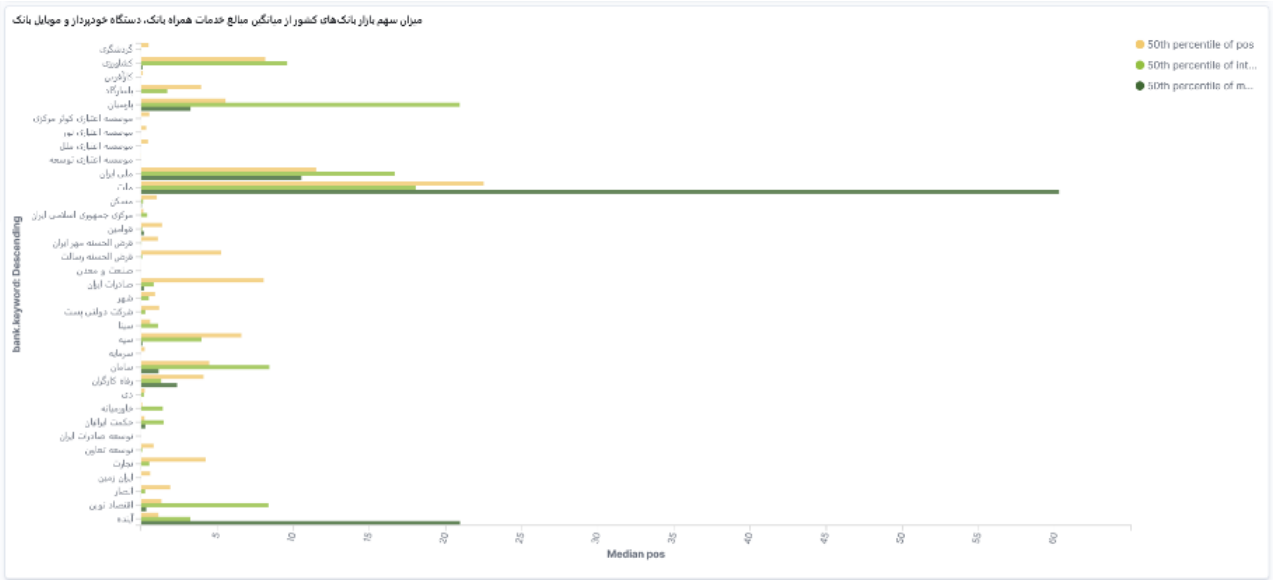
امنیت و کشف تقلب

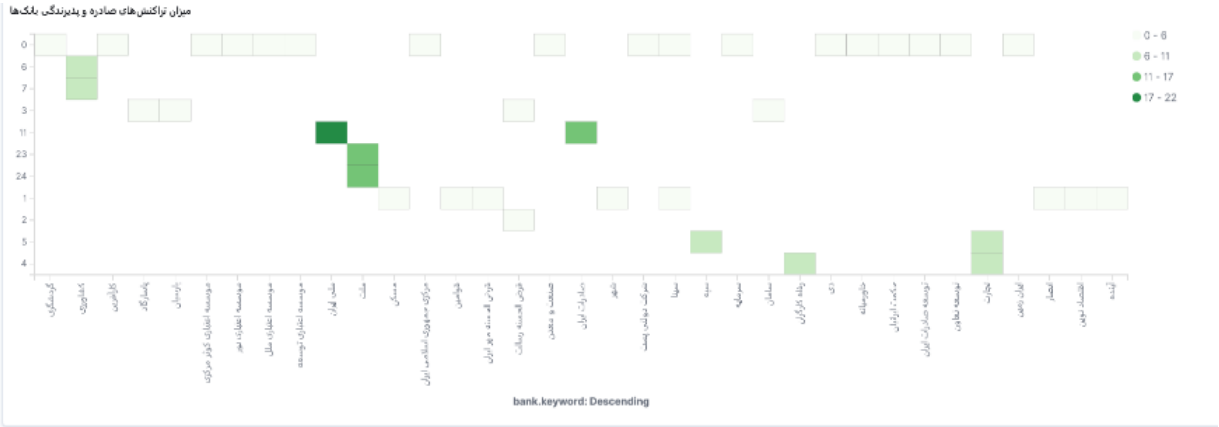
- کشف تقلب
  - تقلب در کارت‌های پرداخت و دستگاه‌های خودپرداز
  - چک
  - کارکنان داخلی
  - کشف پولشویی
  - شرطبندی
- امنیت سایبری

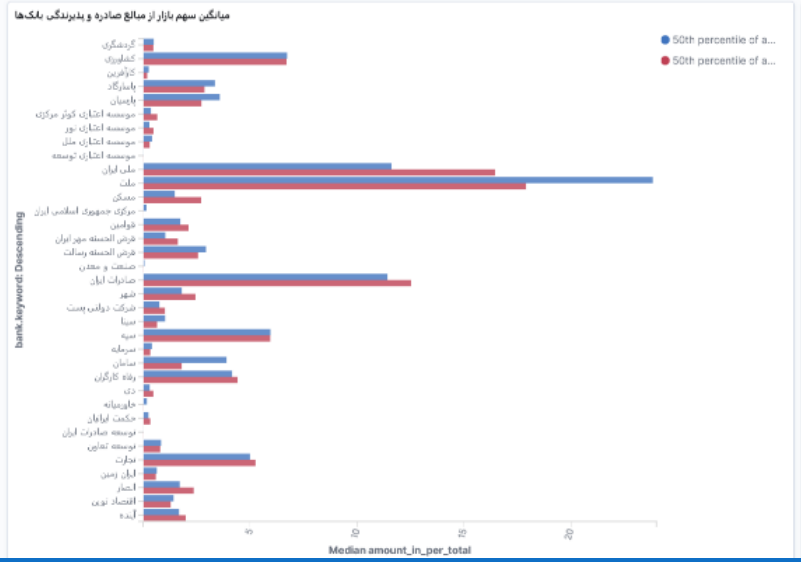
تحلیل‌های درون و برون سازمانی

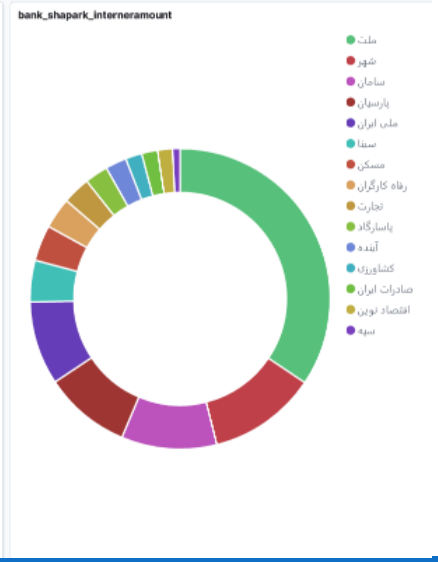
- تحلیل نیروی انسانی
- فرایندکاوی
- خودکارسازی فرایندها
- تحلیل عملکرد و مقایسه بانک‌های رقیب

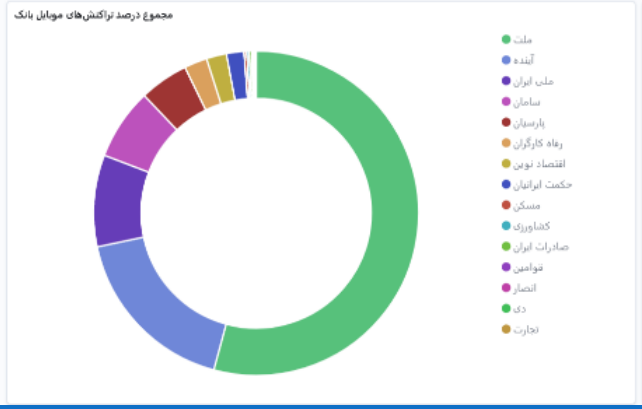
داشبورد تحلیل عملکرد و مقایسه بانک‌های کشور(شاپرک)

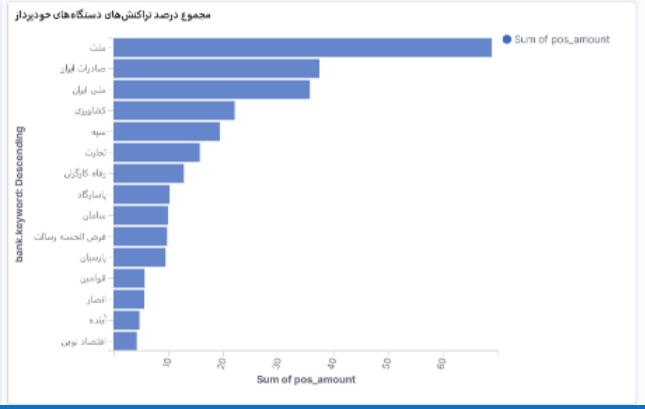




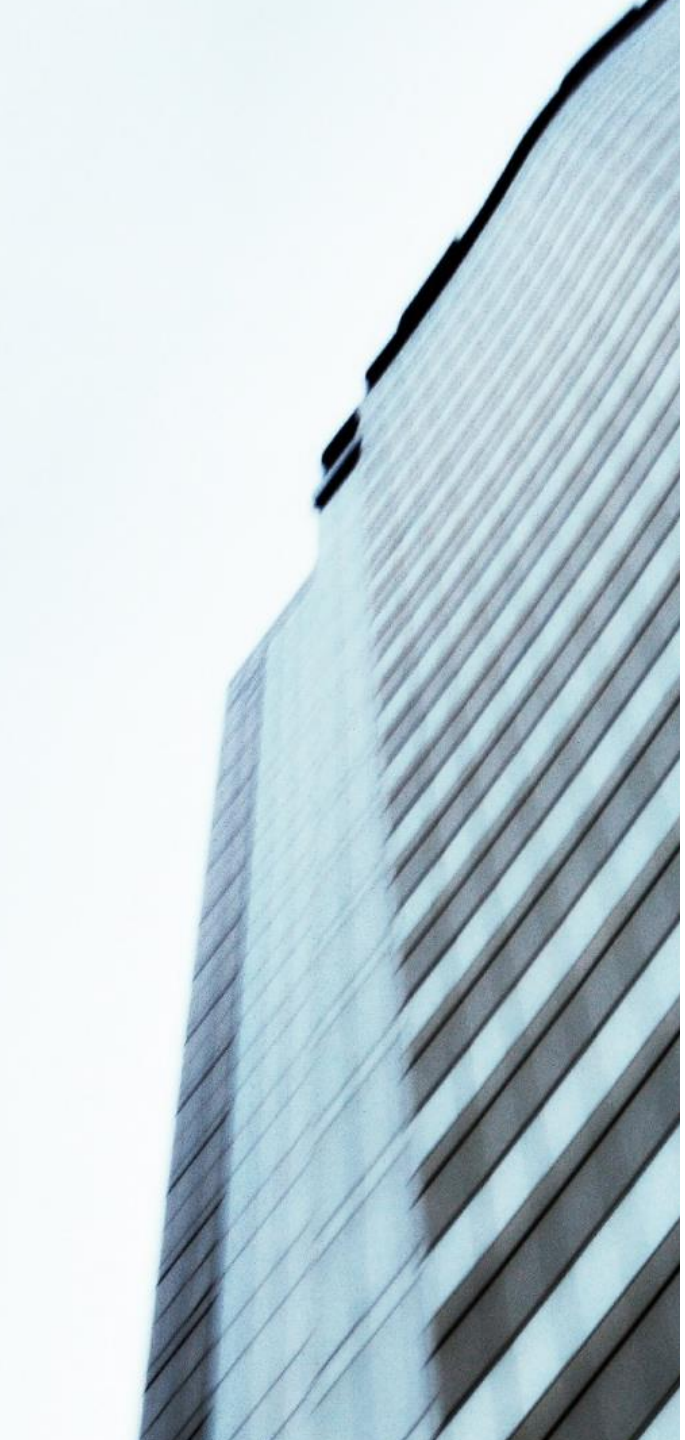








مسیرهای شغلی مرتبط با علوم داده در صنعت بانکی



مسیرهای شغلی مرتبط با علوم داده در صنعت بانکی

- Data Science
- Business Intelligence
- Big Data
- Data Engineering

مهارت‌های کلیدی متخصص علم‌داده

Python

R

SQL

Data Mining Tools

Knime , Rapid Miner,
IBM SPSS Modeler

Excel

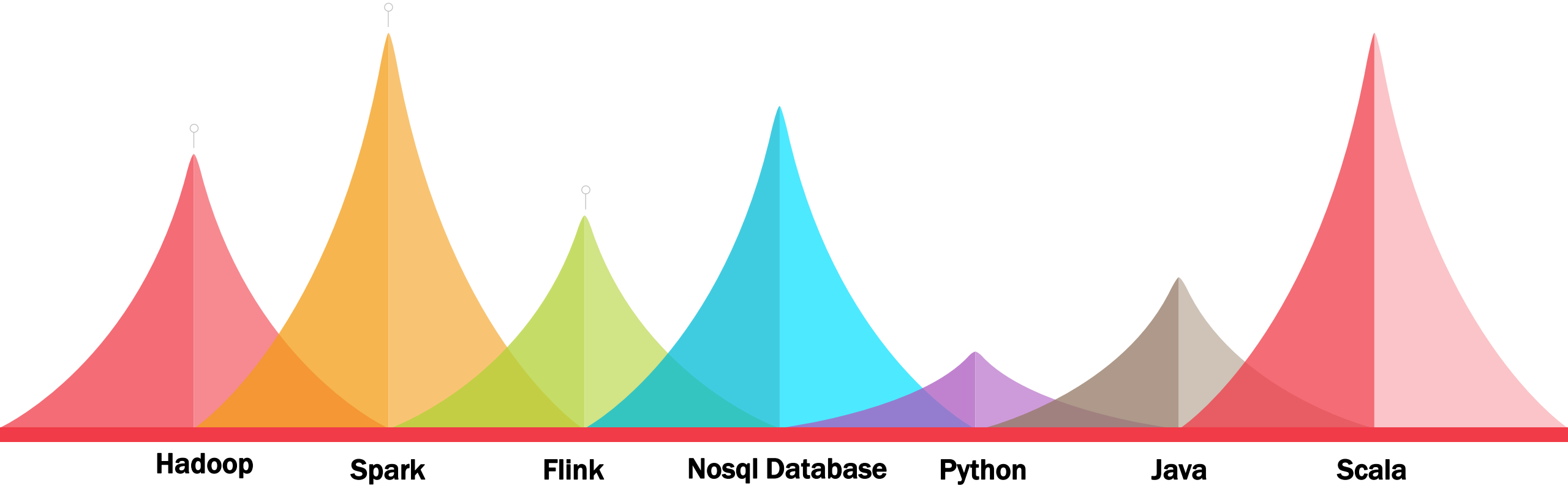
BI Tools

Tableau, Power BI, Qlik

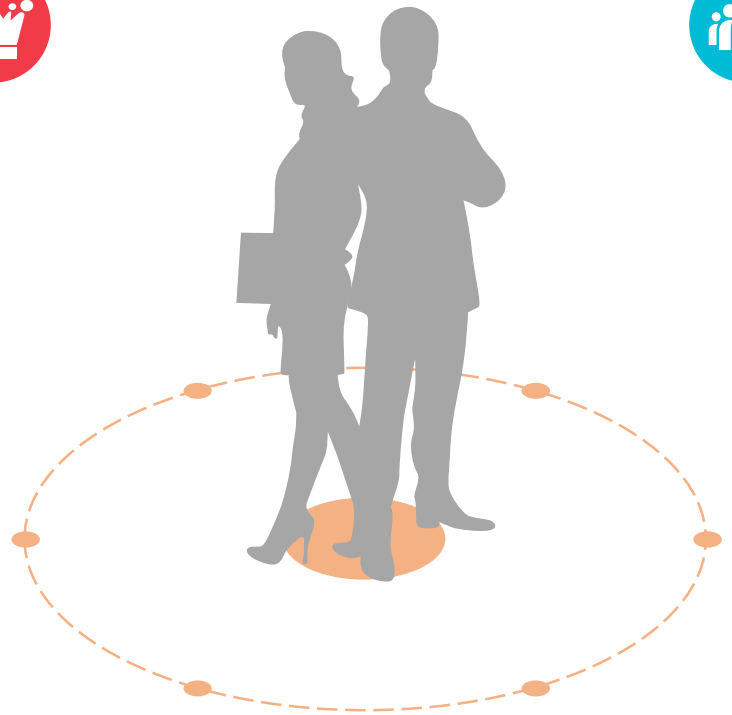
مهارت‌های کلیدی متخصص هوش تجاری

- SQL
- Data Warehousing
- Data Analysis
- Oracle BI
- ETL
- Power BI
- Microsoft BI
- Business Analyst
- Tableau

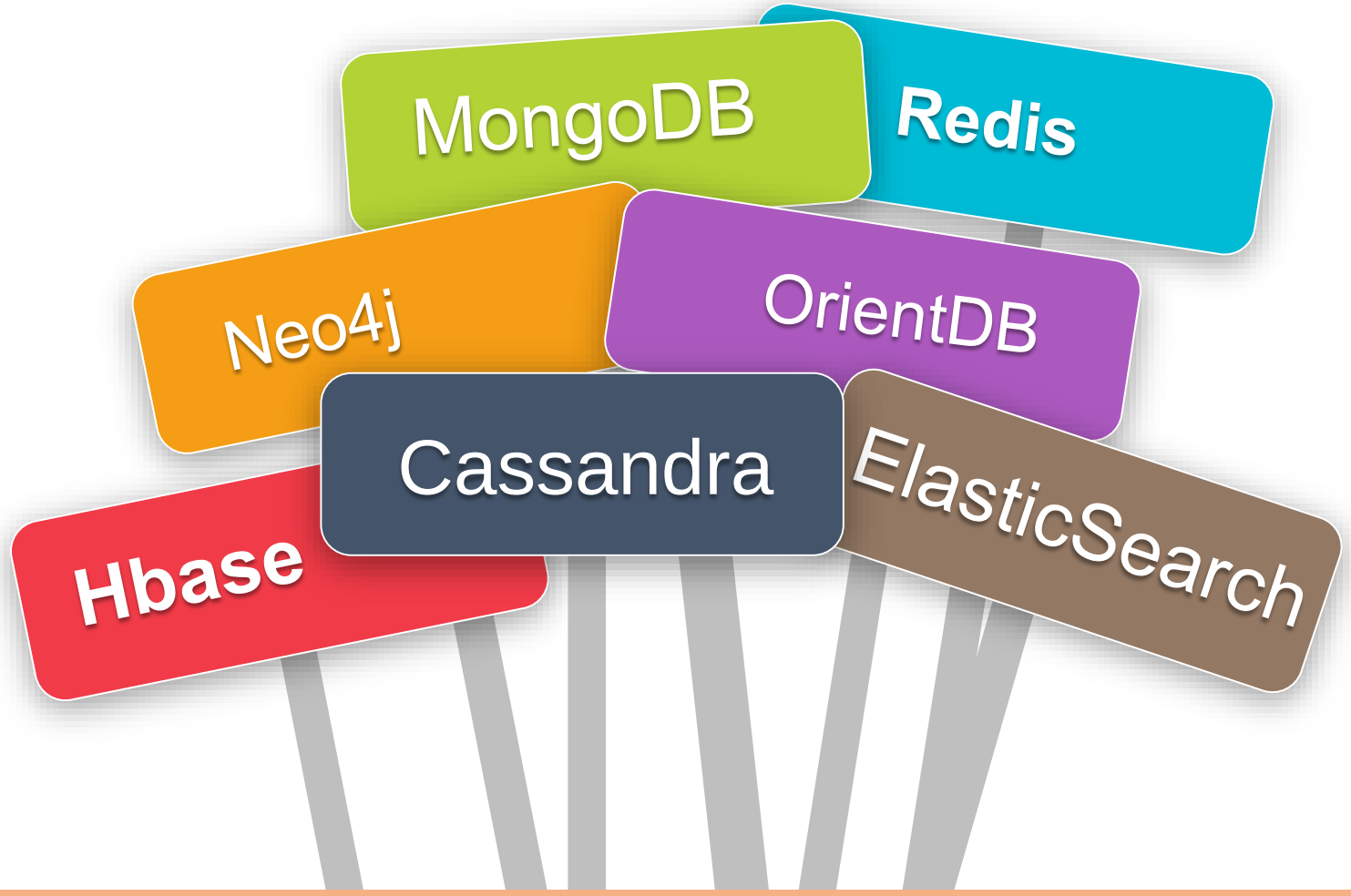
مهارت‌های کلیدی متخصص کلان‌داده



مهارت‌های کلیدی متخصص مهندس داده



- Agile Development
- Linux
- Data Ingestion
- Python / Java / Scala
- Processing Frameworks
- NoSQL & SQL



پرکاربردترین پایگاه‌داده‌های غیررابطه‌ای (NoSQL)

چالش‌های بانک‌های ایرانی در حوزه علوم‌داده

چالش‌های بانک‌های ایرانی در حوزه علوم‌داده

- دیدگاه سنتی در خلق ارزش (فرهنگ تصمیم‌گیری داده‌محور)
- مقاومت سازمانی کارشناسان و مدیران
- جذب استعدادهای داده / تحلیل داده
- کیفیت داده
- تمرکز بیش از حد فنی
- حریم خصوصی، محرمانگی و امنیت داده
- استراتژی و حاکمیت داده
- داده جریانی و لحظه‌ای

طراحی یک برنامه جامع استراتژی داده

تدوین چشم‌انداز، ماموریت‌ها و اهداف بانک در حوزه داده و تحلیل داده در راستای برنامه کلان بانک

تدوین استراتژی و سیاست‌های داده بانک در مواجه با بازار رقابتی طی برنامه‌های یکساله و چهارساله

پیشنهادات بهبود

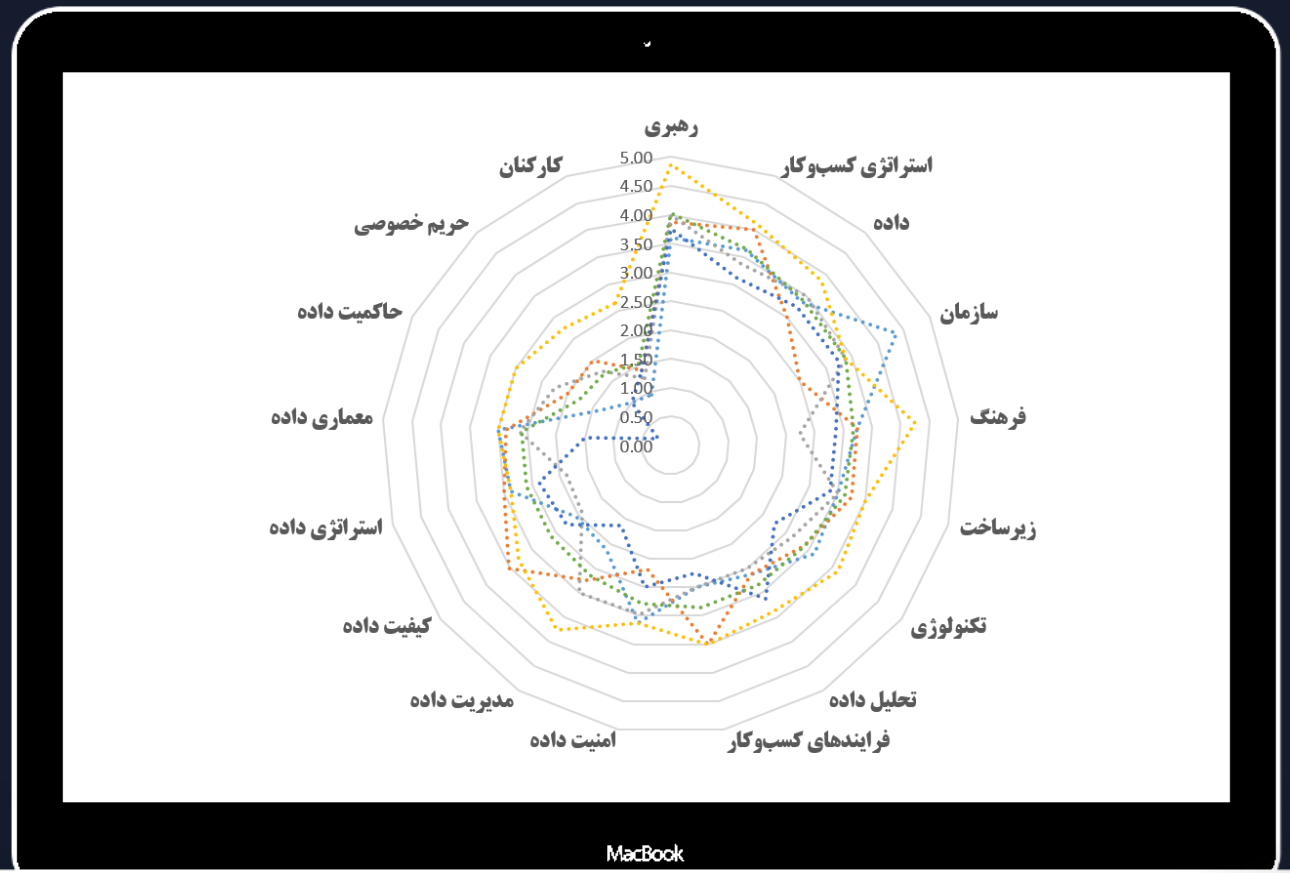
- ارزیابی بلوغ (کلان‌داده، علوم‌داده، مدیریت داده، حاکمیت داده و...)
- طراحی برنامه مدیریت و حاکمیت داده
- تدوین استراتژی‌های حریم خصوصی و اخلاق داده
- طراحی سیاست‌های بانک در جمع آوری و تکمیل منابع داده‌ای
- تدوین برنامه طرح رشد کارکنان و جذب استعدادهای داده / تحلیل‌داده
- طراحی استراتژی داده بانک در محیط اکوسیستم (اشتراکی)
- بررسی طرح‌های اجرایی مبتنی بر کسب درآمد از داده
- شناسایی، امکان‌سنجی و اولویت‌بندی پروژه‌های تخصصی تحلیل‌داده
- تدوین نقشه‌راه استقرار فرهنگ تصمیم‌گیری داده‌محور

معرفی مدل ارزیابی بلوغ علوم داده و کلان داده

- امروزه سازمان‌های اثربخش به دنبال استفاده استراتژیک از داده‌ها همسو با کسب‌وکار خود می‌باشند. داشتن یک استراتژی جامع علوم‌داده و داده‌های عظیم که روح خود را از استراتژی کسب‌وکار می‌گیرد، یکی از الزامات اولیه و حیاتی جهت نیل به موفقیت پایدار در عصر حاضر است.
- استراتژی کسب‌وکار، استراتژی داده، فرایندهای کسب‌وکار، ساختار سازمانی، رهبری، کارکنان، فرهنگ، تحلیل‌داده، حاکمیت داده، حریم خصوصی، امنیت داده، تکنولوژی، زیرساخت، مدیریت داده، معماری داده، کیفیت داده **به همراه مولفه داده** همگی از مولفه‌هایی هستند که برای ارتقا بلوغ علوم داده و داده‌های عظیم حائز اهمیت می‌باشند. در مدل ارزیابی بلوغ علوم‌داده و داده‌های عظیم به تفکیک این ۱۷ مولفه تاثیرگذار مورد ارزیابی بلوغ قرار خواهند گرفت.

نمونه ارزیابی بلوغ علوم داده و داده‌های عظیم

شناخت نقاط قوت، ضعف و گلوگاه‌های کسب‌وکار



Channel Info

Data Science
7,285 subscribers

t.me/dataanalysis
Link

ارتباط با ادمین:
@Datascience
Description

Notifications

VIEW CHANNEL

598 photos

24 videos

69 files

2 audio files

379 shared links

7,285 subscribers

Mohamadreza Mohtat

Principal Data Scientist | Big Data Consultant

Iran · 500+ connections · Contact info

University of Tehran

Allameh Tabataba'i University

Open to · Add profile section · More...

Show recruiters you're open to work — you control who sees this
Get started

Share that you're hiring and attract qualified candidates.
Get started