

موضوع پروژه: تجزیه و تحلیل داده‌های منابع انسانی: تغییر شغل دانشمندان داده

پیش بینی اینکه چه کسانی شغل خود را تغییر می‌دهند.

لطفاً به سوالات زیر پاسخ دهید:

۱- هدف پروژه چیست؟

شرکتی که در زمینه Big Data و Data Science فعال است می‌خواهد دانشمندان داده‌ای را از بین افرادی که برخی از دوره‌های این شرکت را با موفقیت گذرانده‌اند، استخدام کند. بسیاری از افراد برای آموزش خود ثبت نام می‌کنند. شرکت می‌خواهد بداند کدام یک از این داوطلبان واقعاً می‌خواهند بعد از آموزش برای این شرکت کارکنند یا به دنبال شغل جدید دیگری می‌روند. فهمیدن این موضوع به کاهش هزینه و زمان و همانطور به افزایش کیفیت آموزش یا برنامه‌ریزی دوره‌ها و دسته‌بندی داوطلبان کمک می‌کند. اطلاعات مربوط به جمعیت شناسی، تحصیلات و تجربه از ثبت نام نامزدها در دست است. این مجموعه داده، آماده شده برای تحقیقات HR است. همچنین این مجموعه داده به گونه‌ای طراحی شده است که به درک عوامل مؤثر در ترک شغل فعلی توسط افراد کمک می‌کند. کارمند کنونی شرایط یک داوطلب را دارد که یک دوره در شرکت گذرانده است. تغییر شغل کارمند کنونی مانند این است که یک داوطلب یا کاندید در شرکت کار نکند. اطلاعات موجود کمک می‌کند که احتمال اینکه یک کارمند به دنبال شغل جدید باشد یا برای شرکت کار کند را بدست آوریم و همچنین عوامل مؤثر بر تصمیم کارمند را تفسیر کنیم.

در این پروژه تمرکز بر این است که بفهمیم چرا کارمندان شرکت را ترک می‌کنند یا چرا کاندیدها برای کار به شرکت نمی‌پیوندند. با کاوش در این مجموعه داده می‌توان بینش خوبی از مشکلاتی که بخش منابع انسانی در این زمینه با آن سروکار دارد استخراج کرد. از نتایج این مطالعه می‌توان برای افزایش استخدام‌ها از کاندیدهای مناسب استفاده کرد و همچنین عوامل مهم برای تغییر شغل را بیابیم تا کارمندان را از دست ندهیم. با این بررسی عوامل مؤثر در این تغییر شغل‌ها را می‌یابیم.

همچنین در مطالعه داده ما مهم ترین پارامترها و ویژگی‌ها در ماندن کاربران یا استخدام شدن کاندیدها را از میان ویژگی‌ها می‌یابیم. این موضوع به ما کمک می‌کند که بهتر بتوانیم احتمال اینکه یک کارمند به دنبال شغل جدید باشد یا برای شرکت کار کند را پیش بینی کنیم.

۲- منبع داده از کجاست؟

داده یک دیتاست در سایت Kaggle می‌باشد. ولی شرکت تهیه کننده‌ی مجموعه داده مشخص نیست. تنها اطلاعات در مورد شرکت منبع داده‌ها در ادامه آمده است. شرکتی که در زمینه Big Data و Data

Science فعال است می خواهد دانشمندان داده‌ای را از بین افرادی که برخی از دوره های این شرکت را با موفقیت گذرانده‌اند، استخدام کند. بسیاری از افراد برای آموزش خود ثبت نام می کنند. شرکت می‌خواهد بداند کدام یک از این داوطلبان واقعاً می‌خواهند بعد از آموزش برای این شرکت کارکنند یا به دنبال شغل جدید دیگری می‌روند. فهمیدن این موضوع به کاهش هزینه و زمان و همانطور به افزایش کیفیت آموزش یا برنامه‌ریزی دوره‌ها و دسته‌بندی داوطلبان کمک می‌کند. در ادامه آدرس دیتاست آمده است.

۳- چه سوالاتی باید پاسخ داده شوند؟

کدام ویژگی‌ها در باقی ماندن کارمند یا اضافه شدن یک کاندید به شرکت مهم هستند؟ تاثیر هر کدام از ویژگی‌ها در این زمینه چگونه است؟ ارتباط ویژگی‌ها با هم به چه شکل است؟ چگونه null ها را پر کنیم؟ outlier ها چه مواردی را نشان می دهند و آیا می‌توان آنها را حذف کرد؟ اگر می توان این کار را کرد، چگونه به حذف پردازیم تا اطلاعاتی از دست نرود.

4- داده به چه شکلی جمع آوری شده است؟

داده از طریق اطلاعات در دست از کارمندان و کاندیدها از طریق فرم‌های sign up و enrolment که برای ثبت نام هستند، جمع آوری شده است.

5- چه جدول‌هایی در داده وجود دارد؟

جدول داده آموزش، تست و جدول برای نمایش نحوه ارسال پیش بینی‌ها وجود دارد. جدول آموزش دارای اطلاعات افراد به همراه این است که شغل را تغییر داده اند یا در شرکت استخدام شده اند. که ویژگی target این موضوع را نشان می دهد. در جدول تست، تمام اطلاعات افراد وجود دارد و فقط target موجود نیست که باید پیش بینی شود. جدول نحوه ارسال پاسخ در بررسی ما اهمیتی ندارد. داده تست چون لیبل ندارد، باز در بررسی ما جایگاهی ندارد. برای تکمیل اطلاعات برای توزیع و ارتباط ویژگی‌های افراد می توان استفاده کرد ولی در تعیین ارتباط ویژگی با target کارا نیستند.

6- جزئیات جدول‌ها به چه شکلی است؟

داده آموزش یک جدول (۱۴ × ۱۹۱۵۸) است. یعنی اطلاعات ۱۹۱۵۸ فرد را داریم. برای هر فرد ۱۳ ویژگی برای توصیف و یک لیبل وجود دارد. داده تست یک جدول (۱۳ × ۲۱۲۹) است. یعنی اطلاعات



۲۱۲۹ نفر را داریم. در این جدول ۱۲ ویژگی برای هر فرد موجود است ولی لیبل موجود نیست. در ادامه ویژگی‌های ۱۳ گانه با توصیفی کوتاه آمده است.

- enrollee_id : Unique ID for candidate
- city: City code
- city_development_index : Development index of the city (scaled)
- gender: Gender of candidate
- relevent_experience: Relevant experience of candidate
- enrolled_university: Type of University course enrolled if any
- education_level: Education level of candidate
- major_discipline :Education major discipline of candidate
- experience: Candidate total experience in years
- company_size: No of employees in current employer's company
- company_type : Type of current employer
- lastnewjob: Difference in years between previous job and current job
- training_hours: training hours completed
- target: 0 – Not looking for job change, 1 – Looking for a job change

در صورتی که منبع داده پابلیک است لطفا لینک صفحه‌ی داده را بنویسید.

<https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists>

در صورتی که منبع داده پابلیک نیست توضیح دهید:

نکات تحویل

- مهلت ارسال فایل پروپزال تا پایان روز چهارشنبه ۱۷ آذر ماه خواهد بود.



- لطفا فایل خود را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

Proposal _[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:

mojtaba.amiri@ut.ac.ir