禪

برنامهنویسان چگونه میتوانند از هوش مصنوعی بدون ناظر (Driverless AI) استفاده کنند

# خودکارسازی یادگیری ماشین با H2O.ai


خودکارسازی یادگیری ماشینی با
H2O.ai

مشخص شده است یادگیری ماشین و بهویژه یادگیری عمیق در زمانهایی که در دست متخصصان این حوزه قرار گیرد بهمیزان بسیار زیادی مفید واقع میشود و همچنین بهمیزان بسیاری به سخت افزار کامپیوتر وابسته است. با فراگیر شدن GPGPU (واحد پردازش گرافیکی عمومی ) پرقدرت، FPGA (آرایه گیتهای قابل برنامهریزی ) و تراشههای سفارشی مثل Tensor Processing Unit (TPU) گوگل نیز دستیابی به چنین قابلیتهایی آسانتر از گذشته شده است . اما ما همچنان با کمبود متخصصانی از جمله دانشمندان داده و برنامهنویسان یادگیری ماشین مواجه هستیم .

اینکه این کمبود تا چه میزان جدی است به این بستگی دارد که روند آموزش و استخدام دانشمندان داده و توسعهدهندگان توسط سازمانهای مرتبط تا چه حد جدی گرفته میشود. اما هوش مصنوعی بدون ناظر یا Driverless AI قادر است مدلهای یادگیری ماشینی را بدون نیاز به مهارتهای یادگیری ماشینی از طرف کاربران ایجاد کند . تا زمان انتشار این مقاله Driverless AI تنها سیستم یادگیری ماشینی خودکار است که به این شکل عرضه شده است. اما بدون شک ظرف یک سال آینده رقبای جدیدی چه بهصورت مجموعههای منبع باز یا خدمات مستقل رقیب آن خواهند شد.

اگر شرکتها بهجای استخدام دارندگان مدرک دکتری با حداقل 20 سال سابقه کار و ترجیحاً جوان، متخصصان این حوزه از جمله تحلیلگران، برنامهنویسان SQL و برنامهنویسان یادگیری ماشینی را به یک دیگر ارتباط دهند، دستیابی به اهداف مورد نیاز این صنعت سریعتر انجام خواهد شد . علاوه بر این اگر آنها استفاده از ابزارهایی مثل هوش مصنوعی بدون ناظر H2O.ai (Driverless AI) که بخش قابل ملاحظهای از فرآیند یادگیری ماشین را خودکارسازی میکند، در دستور کار خود قرار دهند، میتوانند این گروهها را به طور چشمگیری کارآمدتر کنند . همان طور که مشاهده خواهیم کرد، هوش مصنوعی بدون ناظر یا Driverless AI یک سیستم یادگیری ماشینی است که به طور خودکار هدایت میشود و بدون نیاز به تخصص علم داده قادر به ایجاد و آموزش مدلهای شگفتانگیزی در طی مدت زمان بسیار کوتاه است . با وجود اینکه هوش

[illegible]

## [illegible]



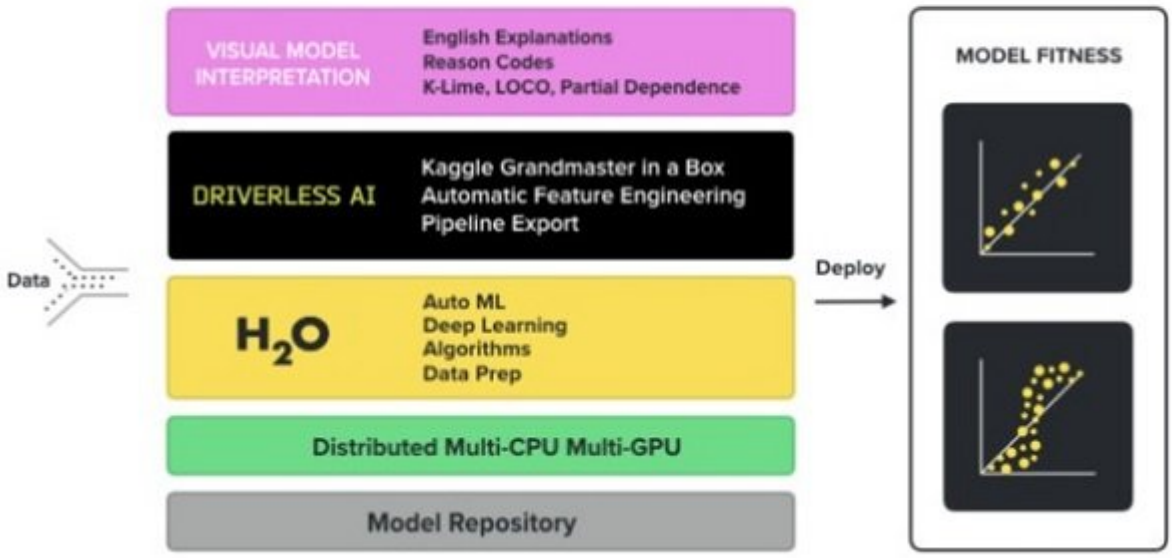
[illegible]

**5G [illegible]**

# [illegible]

[illegible] Auto- [illegible] H2O.ai’s Driverless AI  sklearn, Auto-Weka, Prodigy, Google AutoML, Google Vizier [illegible] H2O.ai [illegible] AutoML [illegible]

[illegible] «[illegible]» [illegible] (1 [illegible])

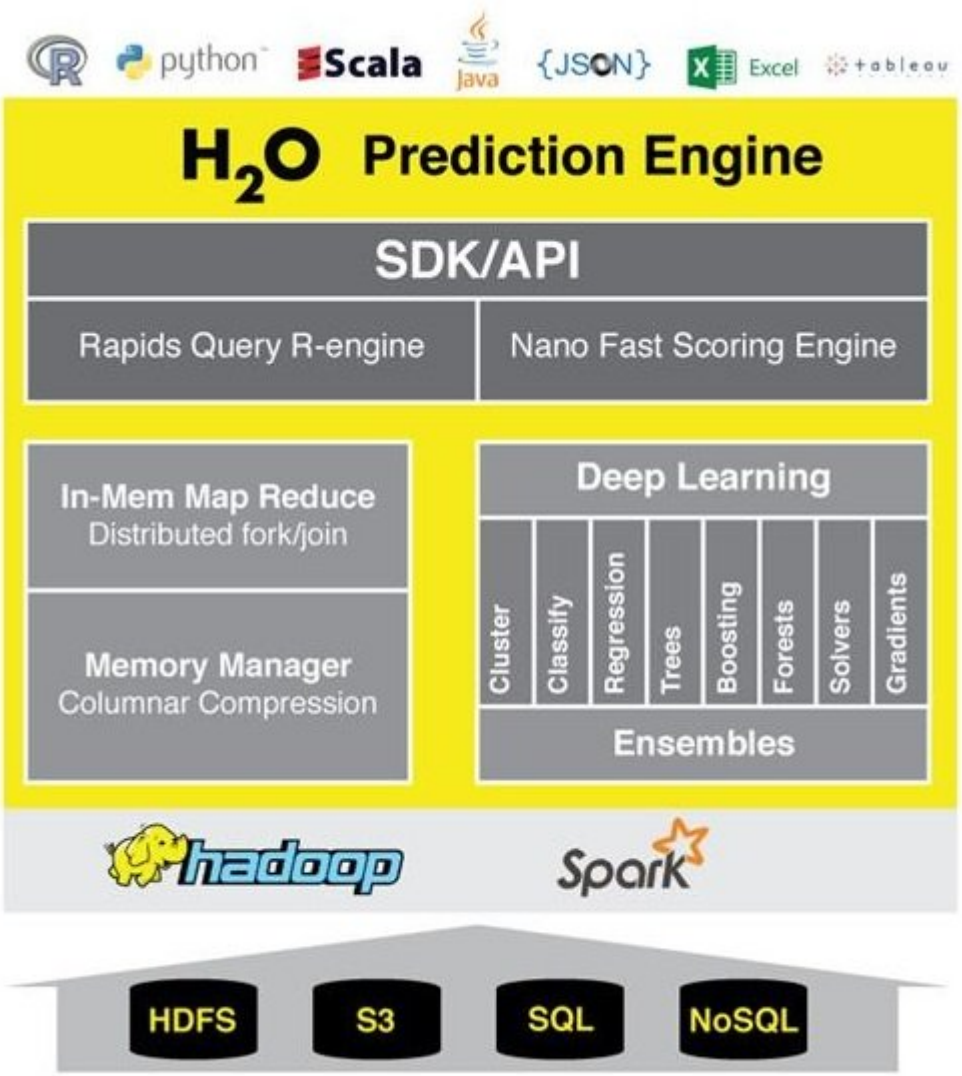


Architecture
VISUAL MODEL INTERPRETATION
English Explanations
Reason Codes
K-Lime, LOCO, Partial Dependence
DRIVERLESS AI
Kaggle Grandmaster in a Box
Automatic Feature Engineering
Pipeline Export
$H_2O$
Auto ML
Deep Learning
Algorithms
Data Prep
Distributed Multi-CPU Multi-GPU
Model Repository
Data
Deploy
MODEL FITNESS

شکل **1**

هوش مصنوعی بدون ناظر یک محصول وب اختصاصی است (نمودار معماری شکل 1 را مشاهده کنید ) که روی پشته H2O.ai (نمودار معماری تصویر پایین شکل 1 را مشاهده کنید ) با هدف تقلید از فرآیندهای مورد استفاده توسط استادان Kaggle برای ایجاد مدلهای عالی ساخته شده است . بخش زیرین پشته H2O.ai نیز بهشکل پیشساخته بهعنوان یک فایل JAR، چرخه پایتون و یک پکیج R موجود است. (شکل 2)

Kaggle یک سایت (https://www.kaggle.com/) ویژه علم داده است که مجموعه دادههای استاندارد را ارائه و مسابقاتی را برای تحلیل آنها برگذار میکند . برخی از این مسابقات دارای پشتیبانی مالی هستند و جوایز قابل توجهی نیز ارائه میکنند (به عنوان مثال برای جایزه اول رقابت چالش الگوریتم غربالگری مسافر TSA مبلغ 500,000 دلار در نظر گرفته شده است ). همچنین، Kaggle دورههای آموزشی و محیطی آنلاین برای علم داده را نیز فراهم کرده است . هر چالش یک سرپرست دارد و تمام کاربران Kaggle توسط همگروهان خود رتبهبندی میشوند. در حال حاضر، در رتبهبندی Kaggle تعداد 95 استاد ارشد و 890 استاد وجود دارد . مطمئناً هزاران استاد دیگر در سطح دانشمندان داده مشغول کار در شرکتهای مختلف هستند که هنوز زمان و شانس رقابت در Kaggle را پیدا نکردهاند . اگر در حوزه علم دادهها مبتدی هستید پیشنهاد میکنم، ابتدا به مسیر https://www.kaggle.com/c/titanic بروید .





python
Scala
Java
{JSON}
Excel
tableau
H2O Prediction Engine
SDK/API
Rapids Query R-engine
Nano Fast Scoring Engine
In-Mem Map Reduce
Distributed fork/join
Memory Manager
Columnar Compression
Deep Learning
Cluster
Classify
Regression
Trees
Boosting
Forests
Solvers
Gradients
Ensembles
hadoop
Spark
HDFS
S3
SQL
NoSQL

شکل **2**

## نصب و تنظیمات H2O.ai Driverless AI

نصب هوش مصنوعی بدون ناظر تنها بهسادگی انتخاب H2O.ai Driverless AI AMI در زمان ایجاد یک نمونه Amazon EC2 است. روی هر کلاود یا ماشین محلی پشتیبانی شدهای شما اصولاً Docker, nvidia-docker و Nvidia driver را نصب میکنید، سپس دایرکتوریهای استاندارد را اضافه میکنید، Driverless AI Docker container را دانلود و نصب و مجوز خود را اضافه میکنید . بعد از اجرا، Driverless AI container یک رابط تحت وب را روی پورت 12345 در اختیار شما میگذارد که میتوانید آن را با مرورگر کروم مشاهده کنید . نمایش پیشفرض (که در شکل 3 مشاهده میکنید ) شامل فهرستی از آزمایشهای شما است.


H2O.ai
1.0.4
Visualize Datasets
Experiments
+ NEW EXPERIMENT

| | TARGET | VALIDATION SCORE | SCORER | ACCURACY | TIME | INTERPRETABILITY | STATUS | TIME |
|---|---|---|---|---|---|---|---|---|
| c65762<br>CreditCard-train.csv | default payment next month | 0.77843 | AUC | 2 | 2 | 5 | Done | 00:03:35 |
| fa764f<br>BNPParibas-train.csv | target | 0.75496 | AUC | 2 | 2 | 5 | Done | 00:08:32 |
| 250823<br>BNPParibas-train.csv | target | 0.75108 | AUC | 1 | 1 | 5 | Done | 00:04:41 |
| 91e471<br>BNPParibas-train.csv | target | 0.76186 | AUC | 5 | 5 | 5 | Running | 00:11:31 |

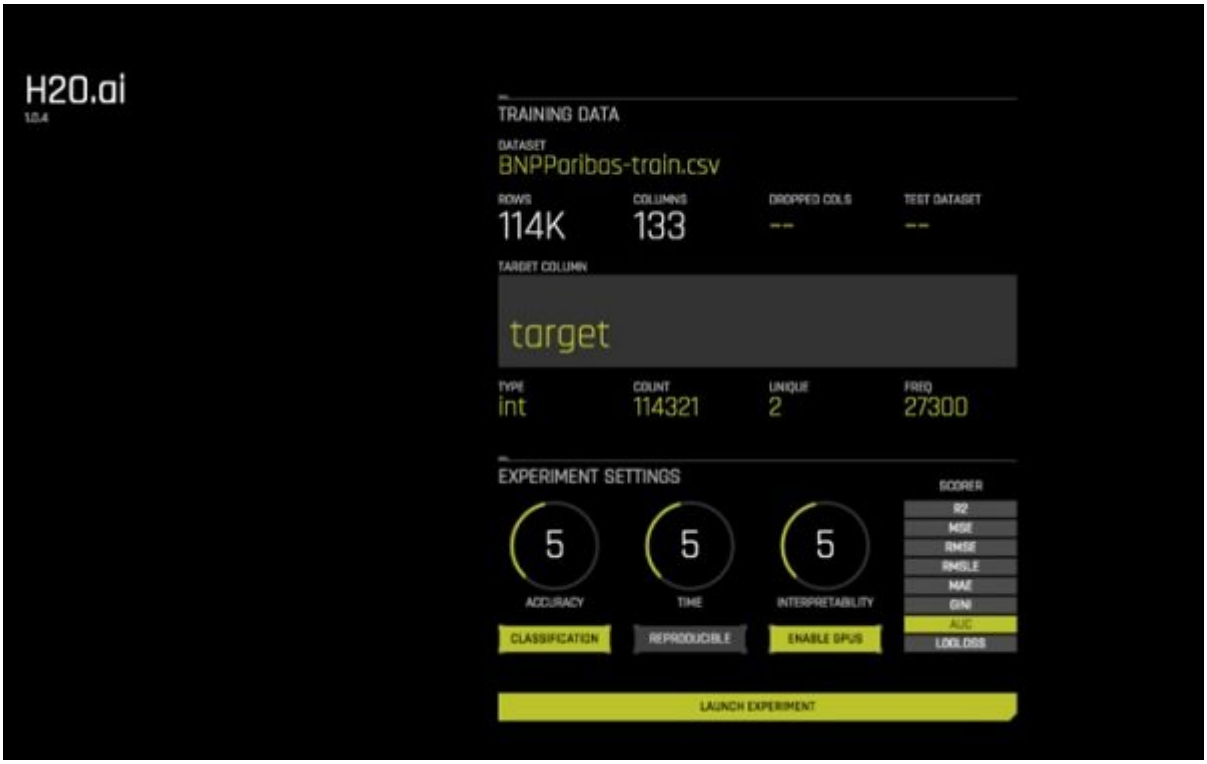
شکل **3**

Driverless AI Docker [illegible] Nvidia driver [illegible] nvidia-docker [illegible] [illegible] GPU [illegible] [illegible] [illegible] Nvidia GPU [illegible] container [illegible] [illegible] Kepler, Maxwell, Pascal [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] Kepler K80 [illegible] [illegible] [illegible] Volta GPU [illegible] Tesla [illegible] [illegible] AWS P2 [illegible] [illegible] Fermi [illegible].

## [illegible]



[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [




**4** ⬚⬚⬚

⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚ Accuracy ⬚⬚⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚: ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚⬚⬚⬚ XGBoost ⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚⬚⬚⬚⬚ ⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚ ⬚⬚⬚ ⬚⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚ ⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚

شده به نوع داده بستگی دارد .
فیلدهای Text ممکن است TF-IDF و ویژگیهای شمارش کلمات را تولید کند . فیلدهای Numeric ممکن است
بهگونهای تبدیل شوند تا مقادیر را به دستههای کوچکتر طبقهبندی کند . (شکل 6)


H2O.ai Experiment 91e471
Show Experiments
TRAINING DATA
DATASET
BNPParibas-train.csv
ROWS
114K
COLUMNS
133
DROPPED COLS
0
TEST DATASET
Yes
TARGET COLUMN
target
TYPE
int
COUNT
114321
UNIQUE
2
FREQ
27300
STATUS: COMPLETE
INTERPRET THIS MODEL
SCORE ON ANOTHER DATASET
DOWNLOAD (HOLDOUT) TRAINING PREDICTIONS
DOWNLOAD TEST PREDICTIONS
DOWNLOAD TRANSFORMED TRAINING DATA
DOWNLOAD TRANSFORMED TEST DATA
DOWNLOAD LOGS
DOWNLOAD SCORING PACKAGE
EXPERIMENT SETTINGS
5
ACCURACY
5
TIME
5
INTERPRETABILITY
CLASSIFICATION
REPRODUCIBLE
ENABLE GPUS
SCORER
R2
MSE
RMSE
RMSLE
MAE
GINI
AUC
LOGLOSS
CPU / MEMORY
Trace
CPU
MEM
ITERATION SCORES - INTERNAL VALIDATION
0.7538
EPOCHS
VARIABLE IMPORTANCE
GPU USAGE
GPU1
GPU2
GPU3
GPU4
GPU5
GPU6
GPU7
GPU8

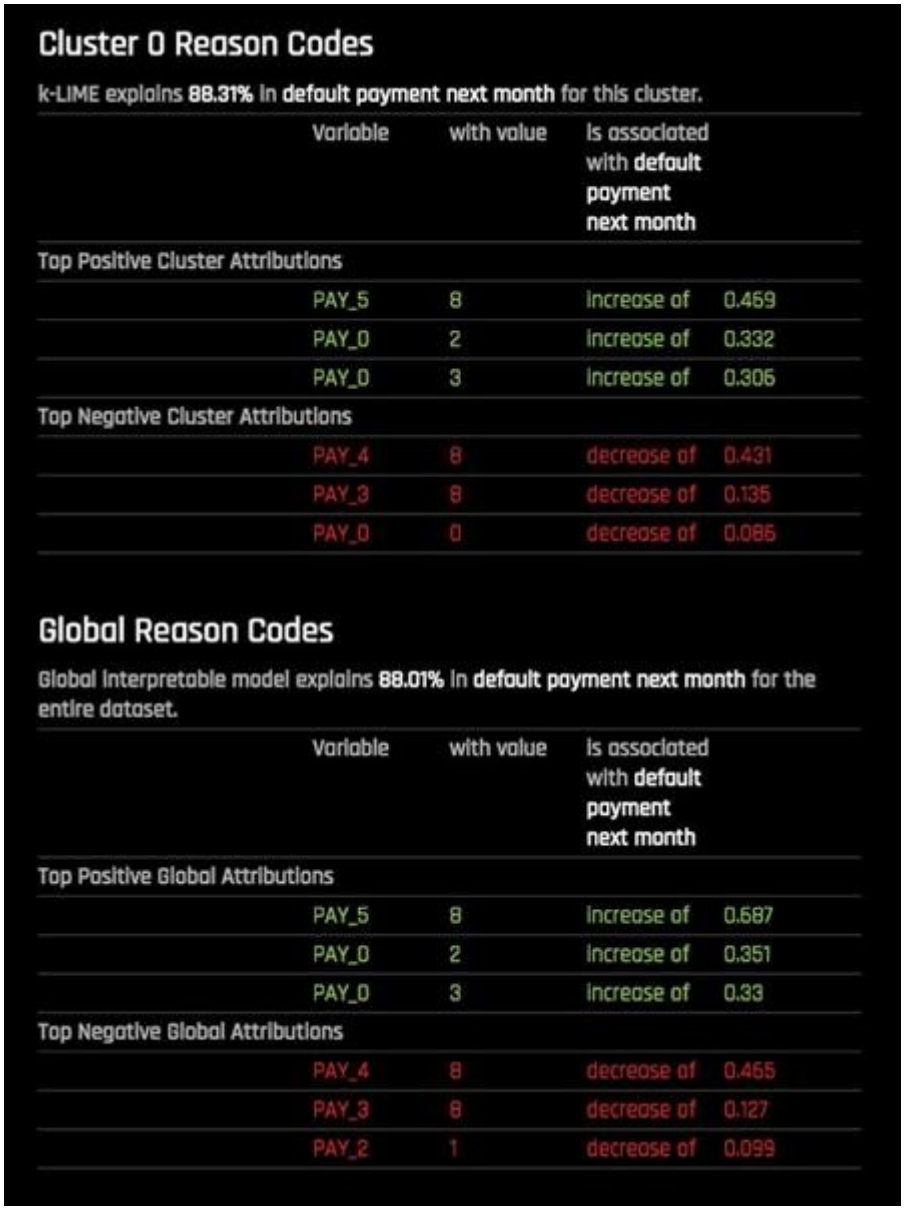
شکل **6**

بعد از اینکه تمام دورهها ارزیابی شدند، Driverless AI یک آموزش کامل را اجرا و با مجموعه ویژگیهای نهایی شروع به تولید پیشبینی میکند . در این مرحله میتوانید تفسیرهای مدل را مشاهده کنید که به محاسبات کمی بیشتری نیاز دارد . (شکل 7)

Single Row Lookup
Column:
H2O Frame Row #
Value:
SEARCH
Plot:
Global
EXPERIMENT
EXPLANATIONS
Global Interpretable Model Explanation Plot
Model Prediction
k-LIME Model Prediction
Actual Target
Ranked Predictions
Variable Importance
Decision Tree
RMSE= 0.108
v50
v56
0.420
0.505
0.633
0.762
0.729
0.893
0.839
0.916
Partial Dependence
v50

شکل 7

صفحه Model Interpretation شامل ترسیمی از طرح توضیحی از مدل تفسیری کلی، نمودار میلهای اهمیت آماری هر متغیر، ساختار درختی مدل جایگزین و ترسیمی از طرح مورد انتظار شرایط موردی است . همه اینها برای کمکِ به تولید توضیحات تقریبی از مدلهای بسیار دقیق با استفاده از تکنیک k-LIME انجام میشود. اساساً چیزی که در اینجا اتفاق میافتد، این است که Driverless AI یک تحلیل k-means را برای تولید کلاسترها اجرا میکند و مدلهای خطی کلاستر را با پیشبینی مدل Driverless AI هماهنگ میکند. از مدلهای محلی برای توضیح ردیفهایی که درون کلاسترهای بزرگ هستند و از مدلهای سراسری برای توضیح ردیفهایی که بیرون کلاسترهای بزرگ قرار دارند استفاده میشود . (شکل 8)



## Cluster 0 Reason Codes

k-LIME explains **88.31%** in **default payment next month** for this cluster.

| | Variable | with value | is associated with **default payment next month** | |
|---|---|---|---|---|
| Top Positive Cluster Attributions | | | | |
| | PAY_5 | 8 | increase of | 0.469 |
| | PAY_0 | 2 | increase of | 0.332 |
| | PAY_0 | 3 | increase of | 0.306 |
| Top Negative Cluster Attributions | | | | |
| | PAY_4 | 8 | decrease of | 0.431 |
| | PAY_3 | 8 | decrease of | 0.135 |
| | PAY_0 | 0 | decrease of | 0.086 |

## Global Reason Codes

Global interpretable model explains **88.01%** in **default payment next month** for the entire dataset.

| | Variable | with value | is associated with **default payment next month** | |
|---|---|---|---|---|
| Top Positive Global Attributions | | | | |
| | PAY_5 | 8 | increase of | 0.687 |
| | PAY_0 | 2 | increase of | 0.351 |
| | PAY_0 | 3 | increase of | 0.33 |
| Top Negative Global Attributions | | | | |
| | PAY_4 | 8 | decrease of | 0.465 |
| | PAY_3 | 8 | decrease of | 0.127 |
| | PAY_2 | 1 | decrease of | 0.099 |

شکل **8**

اگر بخش توضیحات را مشاهده کنید، میتوانید متغیرهایی را که بیشترین میزان مشارکت در مدلهای تفسیری سراسری داشتهاند و مدل کلاسترهای منفرد را ملاحظه کنید . برای مشاهده کلاسترهای مورد نظر خود از منوی کرکرهای Plot استفاده کنید . Driverless AI میتواند برای هرکدام از آزمایشهای یک پکیج رتبهبندی پایتون قابل دانلود تولید کند . شما برای اجرای این پکیجها به ابونتو 16.04 و بالاتر، پایتون 3.6 و مجموعهای از ماژولهای پایتون نیاز دارید . (شکل 9)

H2O.ai
CLUMPY SCATTERPLOTS
CORRELATED SCATTERPLOTS
SPIKEY HISTOGRAMS
SKEWED HISTOGRAMS
BIPLOTS
OUTLIERS
CORRELATION GRAPH
RADAR PLOT
DATA HEATMAP

شکل **9**

اگر روی یک مجموعه داده کلیک کنید، همان طور که در تصویر بالا مشاهده میکنید میتوانید عناصر کلیدی مرتبط با این مجموعه داده را بهشکل دیداری در اختیار داشته باشید . بعد از کلیک روی هرکدام از این عناصر دیداری میتوانید آن را در اندازه کامل مشاهده و دانلود کنید . برای کنترل بیشتر Driverless AI از طریق رابط کاربری وب میتوانید با استفاده از h2oai_client برنامههای کلاینت پایتون را آماده کنید .

مطلب پیشنهادی



زیرساختی برای آینده

مشاغل آینده و مهارتهایی که به آنها نیاز دارید

## ارزیابی یادگیری ماشین خودکار

در مجموع، Driverless AI تحسینبرانگیز است و شما از نتیجه کار آن شگفتزده خواهید شد . این شرکت میگوید استادان Kaggle آن که الگوریتمهای مورد نیاز این سیستم را فراهم میکنند نیز از عملکرد آن غافلگیر شدهاند . مهندسی ویژگی و مدل آموزش اغلب هفتهها زمان میبرد تا به نتیجه قابل قبول دست پیدا کنید . Driverless AI اغلب میتواند ظرف چند دقیقه یا چند ساعت یک پاسخ مناسب را ارائه کند . H2O.ai مدعی است Driverless AI نبوغ استادان ارشد Kaggle را در یک بسته در اختیار شما قرار میدهد . این درست است که متدهای استادان ارشد Kaggle در اختیار شما است، اما دیگرانی که در سازمان شما حضور دارند باید درباره کاری که انجام میدهید اطلاع داشته باشند . در غیر این صورت نمیتوانید درباره وضعیت عملکرد Driverless AI درست تصمیمگیری کنید. بدون برخی پیشزمینههای آماری، مباحث مهندسی ویژگی، کلاسترهای k-means و مدلهای خطی تولید شده برای شما بیمعنا هستند و نتایج بصری نیز شما را به نتیجه مطلوب نمیرساند .

نویسنده :
محسن آقاجانی

منبع :
infoworld
دسته بندی:
کارگاه
تاریخ انتشار :
13:15 - 08/06/1397
برچسب:
Driverless AI - خودکارسازی یادگیری ماشینی - H2O.ai Driverless AI - H2O.ai - یادگیری ماشینی - یادگیری ماشین

---

نشانی
□□□□: https://www.shabakeh-mag.com/workshop/12635/%D8%AE%D9%88%D8%AF%DA%A9%D8%A7%D8%B1%D8%B3%D8%A7%D8%B2%DB%8C-%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D9%85%D8%A7%D8%B4%DB%8C%D9%86-%D8%A8%D8%A7-h2oai