

تشخیص کلاهبرداری از توالی تراکنش‌های بانکی با استفاده از الگوریتم‌های مختلف یادگیری ماشین

کهد آئینی

با نظارت دکتر همت‌یار

محتوای ارائه

۱. تعریف مسئله
۲. آشنایی با مجموعه داده
۳. الگوریتم رگرسیون منطقی
۴. الگوریتم درخت تصمیم
۵. الگوریتم نیوی بیز
۶. پیاده‌سازی و اجرای الگوریتم‌ها
۷. مقایسه‌ی نتایج
۸. نتیجه‌گیری



تشخیص کلاهبرداری و مقایسه‌ی روش‌ها

- هدف تشخیص کلاهبرداری یا سلامت یک مجموعه تراکنش ۲۸ تایی
- حل مسئله با سه الگوریتم مختلف
- بررسی و مقایسه‌ی سه روش با یک دیگر
- تلاش برای یافتن دلایل تفاوت‌ها

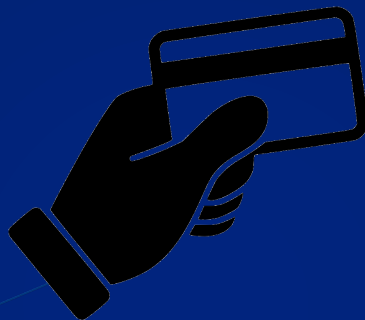
آشنایی با مجموعه داده



dimension
data

ابعاد

۲۸۴۸۰۷x۳۰



۲۸ تراکنش و موجودی حساب
برچسب سلامت توالی
تراکنش‌ها در ستون آخر

۳۰ ستون



تقسیم مجموعه داده

داده‌های
یادگیری

* ۰.۸

$2848 \cdot 8 =$

۲۲۷۸۴۶

داده‌های
آزمون

* ۰.۲

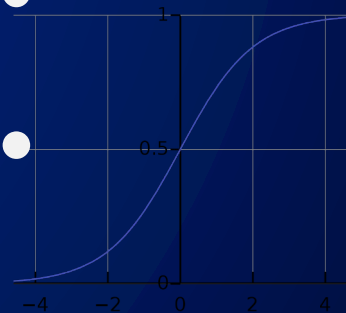
$2848 \cdot 8 =$

۵۶۹۶۱



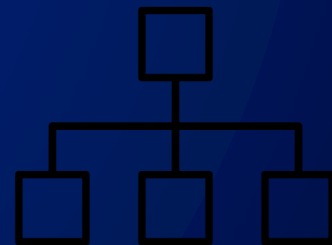
رگرسیون منطقی

- یک روش قوی برای مدلسازی و پیش‌بینی متغیرهای پیوسته
- تحلیل و پیش‌بینی با استفاده از مدل‌سازی احتمالی
- رویکردی موثر برای تحلیل روابط بین متغیرها
- کاربردها و قابلیت‌های این الگوریتم در یادگیری ماشین
- ارائه‌ای جامع برای تحلیل و پیش‌بینی داده‌ها



درخت تصمیم

- از سادگی تا پیچیدگی در تصمیم‌گیری ماشین
- معرفی الگوریتم درخت تصمیم
- یک الگوریتم موثر برای دسته‌بندی داده‌ها
- معرفی الگوریتم درخت تصمیم در یادگیری ماشین
- از جستجوی شاخه‌ها تا تصمیم‌گیری




نیوی بیز

- روشی قدرتمند برای تصمیم‌گیری بر اساس احتمالات شرطی
- دسته‌بندی هوشمند بر اساس تئوری احتمال و آمار
- کاربردها و قابلیت‌های این الگوریتم در یادگیری ماشین
- مدلسازی با الگوریتم نیوی بیز


$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{p(\mathbf{x})}$$








پیاده‌سازی و اجرای الگوریتم‌ها

- پیاده‌سازی با زبان R
- تولید معیارهای ارزش‌سنجی و مقایسه به صورت خودکار

**Fraud-Detection** PublicPinUnwatch 2

main 1 branch 0 tagsGo to fileAdd fileCode

 **kahbodaeni** first commit d016f6d 5 days ago 1 commit

 Dataset	first commit	5 days ago
 Decision Tree	first commit	5 days ago
 Logistic Regression	first commit	5 days ago
 Naive Bayes	first commit	5 days ago
 Pics/Dataset	first commit	5 days ago
 .gitignore	first commit	5 days ago
 پیشنهاد پروژه کارشناسی - کهد آئینی	first commit	5 days ago

مقایسه‌ی نتایج

Parameter	Logistic Regression	Decision Tree	Naive Bayes
Sensitivity	99.98768971	99.96834497	98.28535251
Specificity	58.16326531	78.57142857	85.71428571
Pos Pred Value	99.92793996	99.96307108	99.97495617
Neg Pred Value	89.0625	81.05263158	7.932011331
Precision	99.92793996	99.96307108	99.97495617
Recall	99.98768971	99.96834497	98.28535251
F1	99.95780591	99.96570795	99.12295482
Prevalence	99.82795246	99.82795246	99.82795246
Detection Rate	99.81566335	99.79635189	98.11625498
Detection Prevalence	99.88764242	99.83321922	98.1408332
Balanced Accuracy	79.07547751	89.26988677	91.99981911

Parameter	Best Model	Best Result
Sensitivity	Logistic Regression	99.98768971
Specificity	Naive Bayes	85.71428571
Pos Pred Value	Naive Bayes	99.97495617
Neg Pred Value	Logistic Regression	89.0625
Precision	Naive Bayes	99.97495617
Recall	Logistic Regression	99.98768971
F1	Decision Tree	99.96570795
Prevalence	Logistic Regression	99.82795246
Detection Rate	Logistic Regression	99.81566335
Detection Prevalence	Logistic Regression	99.88764242
Balanced Accuracy	Naive Bayes	91.99981911

نتیجه گیری

- برتری روش نیوی
بیز و رگرسیون
منطقی
- برابری نسبی دو
روش برتر
- یکسان بودن
ماهیت ویژگی ها
دلیل ناکامی
درخت تصمیم



با تشکر فراوان از دکتر همت یار بابت راهنمایی‌های ایشان در انجام پروژه و دکتر صفایی بابت حضور در جلسه

کهد آئینی

kabodaeini@gmail.com

دانشگاه صنعتی شریف

تیر ۱۴۰۲