

MODELS DE CLASSIFICACIÓ

La classificació és una funció de gestió de dades que assigna elements d'una col·lecció a categories, classes o també als anomenats segments. L'objectiu de la classificació és predir amb certesa la classe d'objectiu per a un element donat de la col·lecció. Per exemple, es podria utilitzar un model de classificació per identificar a quina categoria de risc (baix, alt o mitjà) es poden assignar les persones sol·licitants de préstecs. En aquest cas, les persones equivaldrien als elements de la col·lecció (en aquest cas, una població) i els riscos serien les categories.

Un model de classificació comença amb un conjunt de dades de les quals es coneixen les assignacions de les categories. Per exemple, es podria desenvolupar un model de classificació que preveu el risc de crèdit basat en dades observades per a moltes persones sol·licitants de préstecs ja existents, els quals tenen assignat un risc alt, baix o mitjà. A més de la qualificació de crèdit històric, les dades podrien fer un seguiment de l'historial d'ocupació, propietat o lloguer d'habitatges, anys de residència, nombre i tipus d'inversions, etc. La qualificació creditícia seria l'objectiu, els altres atributs serien predictors i les dades de cada client o clienta constituïrien una observació.

La classificació més simple és la classificació binària, en la qual l'atribut només té dos valors possibles; per exemple, qualificació creditícia alta o qualificació creditícia baixa.

En el procés de creació de models, un algoritme de classificació troba relacions entre els valors dels atributs i la categoria o segment on volem col·locar una observació. Diferents algoritmes de classificació utilitzen tècniques diverses per trobar aquestes relacions. Les relacions, doncs, es codifiquen en un determinat model, que serveix després per predir quins elements d'una col·lecció cauran en una categoria o en una altra. Els models de classificació es proven mitjançant la comparació dels valors predits amb els valors objectiu coneguts. Les dades històriques d'un projecte de classificació es divideixen generalment en dos conjunts de dades: un per construir el model; l'altre per provar el model. Suposem que volem predir quins dels nostres clients i clientes podrien augmentar la despesa, si se'ls proporciona una targeta d'afinitat. Podem crear un model mitjançant dades demogràfiques sobre clientela que ha utilitzat anteriorment una targeta d'afinitat. Com que volem predir una resposta positiva o negativa (saber si augmentarà o no la despesa), aquest cas és un clar exemple d'un model de classificació binari.

Les dades de prova han de ser compatibles amb les dades que s'utilitzen per crear el model i s'han de preparar de la mateixa manera que es van preparar les dades de creació.

Normalment, les dades de creació i les dades de prova provenen del mateix conjunt de dades

històriques. Un percentatge dels registres s'utilitza per crear el model; els registres restants s'utilitzen per provar-lo.

Les mètriques de prova s'utilitzen per avaluar la precisió del model de predicció dels valors coneguts. Si el model funciona bé i compleix els requisits comercials, es pot aplicar a dades noves, per predir el futur.

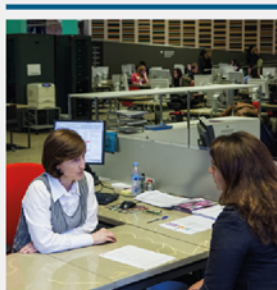
A continuació, anem a descriure els paràmetres que mesuren la qualitat del model emprat:

- La precisió (*accuracy*) fa referència al percentatge de prediccions correctes realitzades pel model, respecte a les classificacions reals de les dades de prova.
- Confusion matrix és una matriu que mostra el nombre de prediccions correctes i incorrectes realitzades pel model, respecte a les classificacions reals de les dades de prova. La matriu és n-per-n, on n és el nombre de classes o segments.
- L'elevació o lift mesura el grau en què les prediccions d'un model de classificació són millors que les prediccions generades aleatòriament. L'elevació només s'aplica a la classificació binària i requereix la designació d'una classe positiva. (Consulteu "Classes positives i negatives"). Si el model en si no té una destinació binària, podem calcular la designació, designant una classe com a positiva i combinant totes les altres classes com una classe negativa.

Aquest paràmetre s'utilitza habitualment per mesurar el rendiment dels models de resposta en aplicacions de màrqueting. L'objectiu d'un model de resposta és identificar segments de la població amb concentracions potencialment altes d'enquestats positius en una campanya de màrqueting. L'elevació revela quina part de la població s'ha de sol·licitar per obtenir el percentatge més alt de possibles persones que responguin.

ROC és una representació gràfica de la sensibilitat davant de l'especificitat per a un sistema de classificació binari, segons varia el llindar de discriminació.

Descobreix tot el que Barcelona Activa pot fer per a tu



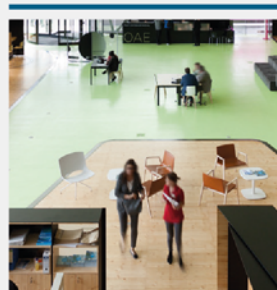
Acompanyament durant tot el procés de recerca de feina

barcelonactiva.cat/treball



Suport per posar en marxa la teva idea de negoci

barcelonactiva.cat/emprenedoria



Serveis a les empreses i iniciatives socioempresarials

barcelonactiva.cat/empreses

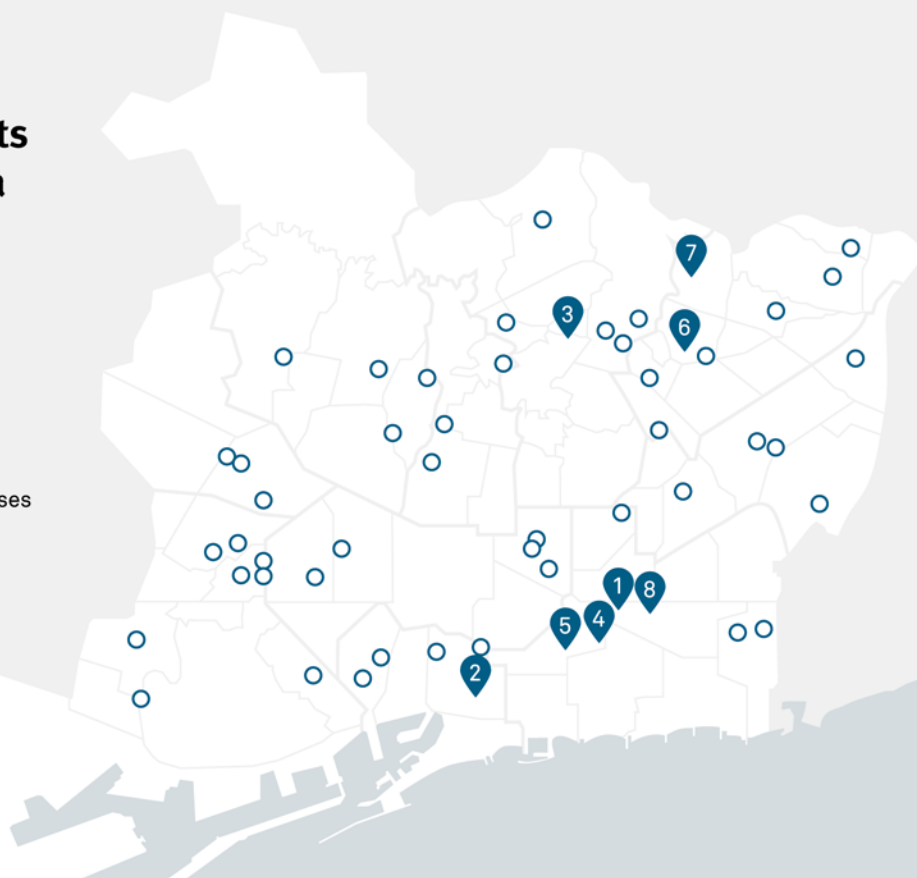


Formació tecnològica i gratuïta per a la ciutadania

barcelonactiva.cat/cibernarium

Xarxa d'equipaments de Barcelona Activa

- 1 Seu Central Barcelona Activa
Porta 22
Centre per a la Iniciativa
Emprenedora Glòries
Incubadora Glòries
- 2 Convent de Sant Agustí
- 3 Ca n'Andalet
- 4 Oficina d'Atenció a les Empreses
Cibernàrium
Incubadora MediaTIC
- 5 Incubadora Almogàvers
- 6 Parc Tecnològic
- 7 Nou Barris Activa
- 8 innoBA
- Punts d'atenció a la ciutat



© Barcelona Activa
Darrera actualització 2019

Cofinançat per:



UNIÓ EUROPEA
Fons Europeu de Desenvolupament Regional

Segueix-nos a les xarxes socials:



barcelonactiva.cat/cibernarium



[barcelonactiva](https://www.facebook.com/barcelonactiva)



[barcelonactiva](https://twitter.com/barcelonactiva)



[company/barcelona-activa](https://www.linkedin.com/company/barcelona-activa)