

Com es decideix a `decidim.barcelona`?

Vicenç Gómez

Premis a projectes innovadors per a la Qualitat Democràtica 2017

1 Resum del projecte (màxim 300 paraules)

Decidim Barcelona és la plataforma de participació ciutadana que l'Ajuntament de Barcelona va posar en marxa al Febrer del 2016. El primer procés ha estat l'elaboració del Pla d'Actuació Municipal¹ i ha donat lloc a un total de 10,860 propostes que han servit per elaborar 1,467 actuacions incloses en el nou pla d'actuació municipal. En aquest projecte es proposa el desenvolupament i l'aplicació de mètodes computacionals i estadístics per extreure coneixement del volum de dades recollides a la plataforma durant el primer període. L'anàlisi té dos objectius principals:

- D'una banda, identificar i corregir el biaix selectiu intrínsec a la plataforma degut a l'escletxa digital. Segons un informe recent, un de cada quatre ciutadans és usuari bàsic o esporàdic d'Internet. Això implica que les dades de participació no són una mostra fiable.
- El segon objectiu consisteix en caracteritzar mitjançant tècniques d'inferència Bayesiana el procés de decisió que ha donat lloc a l'acceptació o rebuig de les propostes de la plataforma. L'anàlisi consistirà en aprendre un model estadístic a partir de les dades associades a cada proposta on es tinguin en compte paràmetres exògens a la plataforma.

Aquests resultats serviran per millorar l'enteniment sobre el comportament col·lectiu en la plataforma, identificar possibles intervencions futures per tal d'incloure sectors de la població infra-representats i en general millorar el rendiment de la plataforma en el futur.

2 Marc conceptual de la proposta

Un dels reptes actuals més importants és com donar sentit als grans volums de dades que la nostra societat genera contínuament. Les xarxes socials, la medicina personalitzada o els sistemes de recomanació, que ja formen part de la nostra vida quotidiana, en són clars exemples.

¹<https://decidim.barcelona/processes/pam>

Aquest projecte considera aquest repte en el marc de la democràcia directa. En Febrer del 2016, l'Ajuntament de Barcelona va encetar la plataforma **decidim.barcelona**. Aquesta plataforma ha donat cabuda als processos de participació ciutadana elaborats durant el mandat. És una plataforma construïda amb eines lliures, similar a la plataforma **decide.madrid**, desenvolupada per l'Ajuntament de Madrid.

El procés participatiu iniciat pel Ajuntamentt ha tingut com a objectiu fer els ciutadans partíceps d'un procés on tinguessin la possibilitat d'evaluar i discutir les propostes realitzades pel govern, així com oferir la possibilitat de crear de noves. Durant l'elaboració del pla estratègic, la plataforma **decidim.barcelona** ha permès registrar, visualitzar i interaccionar amb les propostes institucionals i ciutadanes.

El procés de participació s'ha desenvolupat tant de manera presencial mitjançant reunions com de manera online en la plataforma digital. Les reunions s'han caracteritzat per l'esforç de fer arribar la participació a tots els diferents àmbits territorials de la ciutat, donant lloc a 410 reunions presencials entre Gener i Abril de 2016, on van assistir un total de 11,577 persones i 2,099 organitzacions. Es calcula que la participació presencial durant la primera etapa ha estat aproximadament d'un 43%. La resta de participació, el 57% aproximadament s'ha donat a través de la plataforma online **decidim.barcelona**.

Els resultats d'una primera etapa de participació són 10,859 propostes que han rebut 165,087 suports i 18,191 comentaris.

De moment, han donat lloc a la definició de possibles actuacions, incloses en el Pla d'Actuació Municipal. El procés de definició de les actuacions ha estat portat a terme per l'equip de govern i ha inclòs la revisió, agrupació i reelaboració de les propostes provinents de la ciutadania (organitzada i no organitzada) juntament amb les cites presencials i de l'Ajuntament. Les actuacions s'han definit tenint en compte els continguts de les propostes, el suport que han rebut, els comentaris realitzats i les cites presencials associades. Aquestes actuacions tenen relacionades una o més propostes i cada proposta acceptada està inclosa en una (o més) actuaciones del pla.

Com a exemple, la proposta acceptada amb més suports (1,720) ha estat el *Cubrimient de la Ronda de Dalt al seu pas per la Vall d'Hebrón*. Aquesta proposta ha quedat inclosa en diferents actuacions: (1) *Processos participatius per a la transformació de lavinguda Meridiana i la ronda de Dalt*, (2) *Ronda de Dalt als barris d'Horta-Guinardó* i (3) *Transformar la ronda de Dalt*. Al mateix temps, l'actuació (1) inclou quatre propostes més relacionades (possiblement similars) a la proposta del cobriment de la ronda de dals, com per exemple, *Instal·lació fixa d'estació control contaminació ambiental a la Ronda de Dalt/Av. Meridiana*. Per les propostes rebutjades es dona una breu explicació on s'explica possibles motius pels quals la proposta no s'ha acceptat.

Aquestes dades representen una oportunitat única per analitzar les demandes i interessos col·lectius a Barcelona. La motivació d'aquesta proposta és precisament desenvolupar eines analítiques que se suportin en el rigor científic per analitzar i fer inferències estadístiques.

Perfil No usuari

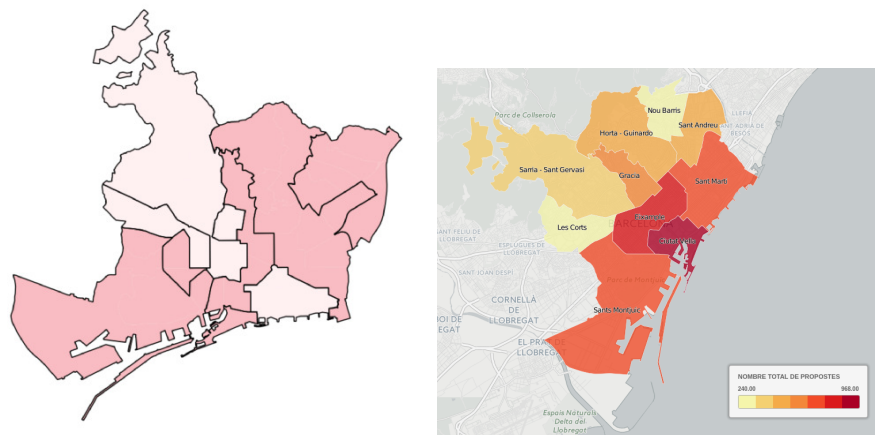


Figure 1: L’escletxa digital i la participació per districte durant l’etapa de participació ciutadana. A l’esquerra, el perfil digital *no connectat* distribuït per nivell de renda segons [5]. A la dreta, un mapa on s’observa el nombre de propostes per districte. La primera part d’aquest projecte proposa caracteritzar (i corregir en la mesura del possible) la influència de l’escletxa digital a Barcelona en la plataforma `decidim.barcelona` utilitzant dades de diferents fonts.

3 Explicació del projecte

A continuació es detallen les dues parts d’aquest projecte. Per cada part es descriu el problema considerat, la metodologia per solucionar-ho i les accions a desenvolupar.

3.1 L’escletxa digital a `decidim.barcelona`

Un dels principals problemes en estudis analítics sociològics i/o polítics és l’existència de biaix selectiu de mostra (en anglès *sample selection bias*), que està associat amb el procés de recollida de dades. Aquest tipus de biaix està induït per la selecció preferent d’unitats particulars de dades, normalment determinades per factors desconeguts, i representa l’obstacle més gran per poder fer vàlides segons quines inferències estadístiques i de causalitat. El biaix selectiu es un de diferents tipus de biaix que existeixen a les dades [1].

Per exemple, si es considera un estudi típic per analitzar l’efecte d’un programa d’entrenament en guanys monetaris, és sabut que els individus que tenen salaris més elevats tendeixen a informar dels seus salaris més sovint que els individus menor reenumerats, els quals informen menys freqüentment, donant lloc a inferències esbiaixades.

En el context d'aquest projecte aquest biaix selectiu està principalment capturat per l'anomenada escletxa digital. *L'escletxa digital fa referència a la desigualtat entre les persones que poden tenir accés o coneixement en relació a les noves tecnologies i les que no* [5]. L'informe [5] ha posat de manifest l'existència d'aquest problema a la ciutat de Barcelona. Entre altres resultats, es posa de manifest que un de cada quatre ciutadans és usuari bàsic o esporàdic d'Internet. Algunes variables claus per determinar el tipus de perfil d'usuari d'Internet són l'edat i el nivell d'estudis. *Tenir una edat avançada, amb un baix nivell d'estudis, treballar en tasques de la llar o ser jubilat i viure en barris de renda baixa augmenten les probabilitats de ser usuari bàsic, esporàdic o directament no ser usuari d'Internet.* Des d'aquest punt de vista, el procés de recollida de dades online a la plataforma `decidim.barcelona` està afectat per un biaix selectiu degut a l'escletxa digital, on sectors significants de la població no hi són representats i d'altres en són sobre-representats. La Figura 1 (esquerra) mostra un exemple del perfil no connectat distribuït per nivell de renda [5]. A l'esquerra, el perfil digital *no connectat* distribuït per nivell de renda. A la dreta, un mapa on s'observa el nombre de propostes per districte ².

El primer objectiu d'aquest projecte és corregir i contrastar el biaix selectiu produït en el procés de participació a `decidim.barcelona`.

Com s'ha comentat anteriorment, és important destacar l'existència d'un procés participatiu *offline*, caracteritzat per l'esforç de fer arribar les reunions presencials a tots els diferents àmbits territorials de la ciutat. A la dreta de la Figura 1 s'observa la representació participativa dels diferents districtes en forma d'activitats presencials. Aquest procés, si bé no resta exempt de possible biaix selectiu, no està tant afectat per l'escletxa digital i suposa efectivament una eina per corregir i contrastar el biaix produït en el procés *online*, que és un dels objectius d'aquest projecte.

Existeixen diferents tècniques i metodologies per tractar de corregir el biaix selectiu [4, 3, 2]. La tècnica més utilitzada consisteix en realitzar una reponderació de les dades observades. Aquestes correccions es realitzen a partir d'una estimació més objectiva que no estigui influenciada per mateix biaix. En el marc d'aquest projecte, és realista pensar que aquest mètode és aplicable, ja que existeixen moltes dades poblacionals digitalitzades, com per exemple, les dades del cens electoral o del padró municipal. A més a més, es compta amb l'avantatge del procés de reunions presencials que ha evolucionat en paral·lel al procés *online*, suposant una font de dades molt valuable per realitzar aquest tipus de correccions.

La metodologia principal, doncs, consistirà en el creuament de diferents bases de dades per tal de calcular la corresponent reponderació estadística i poder aplicar correccions a les estimacions de participació ciutadana actuals.

Les accions a desenvolupar en aquest anàlisi seran les següents:

- 1.1.-** Caracterització quantitativa de l'impacte de l'escletxa digital en la plataforma digital `decidim.barcelona`. Aquesta acció requereix el creuament de dades provinents de la plataforma `decidim.barcelona` (processos online

²https://matteodecidimbcn.carto.com/viz/3f484bd0-4cd7-11e6-9a4f-0ee66e2c9693/public_map

i presencials) com altres possibles dades no esbiaixades (cens electoral o padró municipal, per exemple), així com les dades de l'informe [5].

- 1.2.- Correccions de les mesures actuals de participació ciutadana que no contemplen aquest tipus de biaix selectiu. Aquesta acció donarà lloc a una reponderació de les mesures de participació juntament amb una estimació del la incertesa estadística (en forma d'interval de confiança) per aquelles participacions infra-representades.
- 1.3.- Identificació possibles usos futurs d'aquestes tècniques implementades. Per exemple, servir de suport a la presa de decisions per tal de fomentar l'ús de tecnologies digitals en contextes particulars, o el disseny de mecanismes a implementar en la mateixa plataforma per tal d'associar un indicador quantitatiu de (sobre)-representació de determinats usuaris o col·lectius en detriment d'altres.

3.2 Decisions sobre les propostes i creació d'actuacions

La segona part d'aquest projecte considera el procés de decisió mitjançant el qual les propostes s'han acceptat o rebutjat. Aquest procés ha estat portat a terme per l'equip de govern i ha consistit en la revisió, agrupació i reelaboració de les propostes provinents de la ciutadania (organitzada i no organitzada) juntament amb les cites presencials i de l'Ajuntament.

El segon objectiu d'aquest projecte és extreure el coneixement automàticament de les dades de la plataforma decidim.barcelona per tal d'entendre millor el mecanisme de decisions sobre l'acceptació/rebuig de les propostes. L'enfoc es basa en modelar explícitament el procés que ha assignat a cada proposta una decisió utilitzant mètodes analítics basats en dades [6]. Concretament, es formalitzarà el problema d'extracció de coneixement com un problema d'aprenentatge automàtic semi-supervisat amb estructura. En aquesta formulació, cada proposta té associades un conjunt de característiques descriptives o predictors. Per exemple, el número de vots, l'eix en el qual està definida, el contingut (text) de la proposta, la localització (districte), etc. En general, qualsevol dada associada a una proposta que hagi estat mesurada. Aquestes variables es poden representar mitjançant una matriu $n \times p$ de dades, on n és el número de propostes (10,859) i p és el número de descriptors (encara per determinar). La Figura 2 (blau, esquerra) pretén il·lustrar aquesta representació.

D'altra banda existeixen les variables resposta. En el nostre cas definides com la decisió d'acceptar o rebutjar una proposta, juntament amb quines actuacions s'han relacionat i el text argumental en cas que la proposta fou rebutjada. Aquesta estructura particular en les variables resposta (classificació binària juntament amb una estructura complexa d'interaccions amb les actuacions així com l'existència de llenguatge natural) representa una oportunitat i un problema interessant que no es pot adreçar mitjançant mètodes tradicionals d'aprenentatge automàtic. Aquestes apareixen a la Figura 2 a la dreta, com a *decisions*, en verd o vermell.

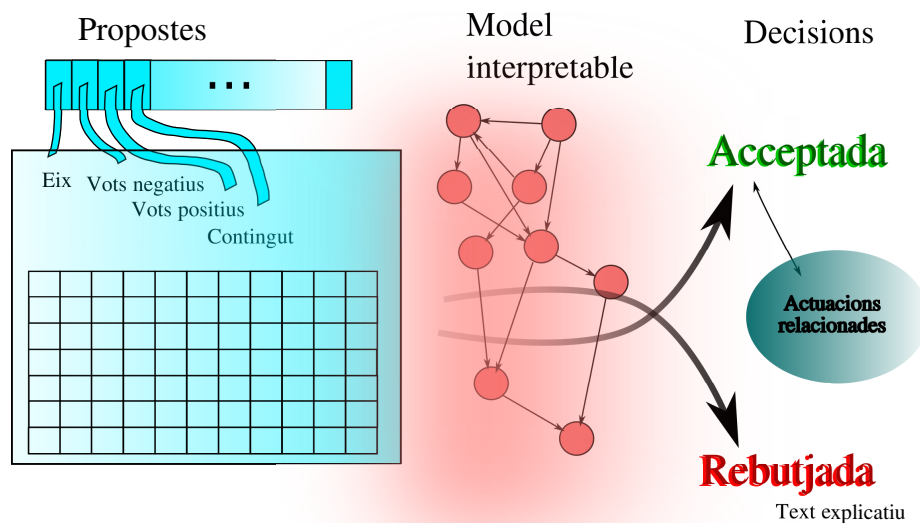


Figure 2: Associació entre propostes i decisions. La segona part del projecte consisteix en construir un model matemàtic (en vermell, part central) a partir de les dades visibles a la plataforma `decidim.barcelona` (descriptors de les propostes, en blau a la dreta) per poder extreure automàticament el coneixement interpretable del procés de que ha donat lloc a les decisions d'acceptació o rebuig de les propostes (part dreta de la figura).

La formalització d'aquest problema en termes matemàtics inclou un model predictiu que és capaç de transformar els predictors en variables resposta. És aquest model predictiu l'objecte a desenvolupar en aquesta part del projecte. El model conté uns paràmetres (o graus de llibertat) els valors dels quals s'estableixen resolent un problema d'optimització matemàtica. Típicament es minimitza l'error entre les variables resposta i les prediccions del model a partir de les variables predictores.

És important que el model predictiu sigui interpretable, és a dir, sigui avinent per fer extracció de coneixement. És bastant freqüent en aquest tipus d'anàlisi utilitzar models de caixa negra (*black box models*), on la interpretabilitat del model queda sacrificada per tal d'obtenir més poder predictiu. En aquest projecte aquesta limitació s'adreçarà fent servir models senzills, on la solució del problema sigui única i interpretable. Hi ha diverses tècniques per aconseguir-ho, totes passen per forçar estructura en la solució del problema que, de manera agnòstica, resulti en un problema amb una única solució òptima global.

Una possible limitació important per assolir aquest segon objectiu és la presència d'un altre tipus de biaix, diferent al biaix de selecció mencionat en

l'apartat anterior. En aquest cas, produït per l'existència de variables latents, les quals no han estat mesurades. Aquestes variables, doncs, no formen part del conjunt de predictors, però poden tenir una influència significativa en el procés per determinar si una proposta ha estat acceptada o no. Per exemple, en el cas que ens ocupa, variables latents poden representar criteris de decisió que puguin haver estat contemplats durant la revisió, agrupació i reelaboració de les propostes i que no apareguin explícits en cap lloc de la plataforma.

L'existència d'aquestes variables *frustrades* (en anglès *confounders*) provoca que el model no sigui indetectable, que vol dir que no es puguin traçar relacions de causalitat entre les variables predictores i les variables resposta [7]. No obstant, això no impedeix poder caracteritzar i establir correlacions i dependències estadístiques robustes que serveixin per identificar les variables més rellevants alhora de predir si una proposta ha estat acceptada o no.

La metodologia principal, doncs, d'aquesta part consistirà en l'aprenentatge automàtic d'un model analític a partir de les dades associades a cada proposta i resultats del pla d'actuació que permeti caracteritzar el procés de decisió a la plataforma.

Les accions a desenvolupar en aquesta segona part seran les següents:

- 2.1.-** Anàlisi quantitatiu de la interacció de les diferents característiques descriptives de les propostes i la relació d'interaccions entre les diferents propostes acceptades i actuacions definides. Aquesta part és important per assegurar que les dades d'entrada són adequades al model utilitzat.
- 2.2.-** Definició i optimització d'un model matemàtic que capturi el procés de decisió i que sigui interpretable. Aquest model identificarà les característiques de les propostes que estadísticament siguin rellevants per poder concloure el resultat d'acceptació o rebuig d'aquesta.
- 2.3.-** Identificació de possibles usos futurs d'aquestes tècniques implementades. Per exemple, la construcció d'un sistema de recomanació automàtic on, d'acord a diferents criteris, que suporti la presa de decisions de l'equip de l'Ajuntament.

4 Públic objectiu de la recerca i beneficis potencials generats dels resultats

Aquest projecte clarament reconeix les pràctiques de participació que són iniciativa de la ciutadania i aprofundeix en els mecanismes de democràcia i de participació de l'Ajuntament de Barcelona. Els resultats no només són interès científic tant en l'àmbit socio-polític com el computacional i estadístic, sino que, a més a més, generaran una utilitat directa en forma d'eines implementades directament a la plataforma. Per tant, aquest projecte també ajuda a avançar en la coproducció de polítiques públiques.

La tasca de quantificar la influència de l'escletxa digital en `decidim.barcelona`, l'eina més important de democràcia participativa del govern, és imprescindible

per garantir un grau de participació de persones i col·lectius en risc de ser invisibilitzats en el disseny i execució de les polítiques participatives actuals, tenint en compte diversitat d'origen, diversitat funcional, paritat per sexe i altres. El projecte també aprofundeix en la col·laboració entre Ajuntament i el teixit associatiu pel que fa a la gestió d'allò públic.

El model desenvolupat a la segona part del projecte és rellevant per diferents raons. Primera, generar models matemàtics que permetin explicar de manera objectiva quins valors influeixen en les valoracions de l'equip de govern facilita la transparència. Segona, adreça un dels reptes actuals més importants de donar sentit als grans volums de dades que la nostra societat genera. Important, el model proposat és interpretable, validable i també modificable. Des d'aquest punt de vista, la presa de decisions i l'automatització van lligades i depenen del factor humà, evitant possibles efectes de discriminació algorítmica.

5 Pla de treball

El projecte serà coordinat pel sol·licitant que realitzarà les tasques de supervisió i seguiment continuat. Donat que el grau d'interdependència entre les dues parts del projecte es baix, aquest es pot entendre com dos sub-projectes relativament independents. Es proposaran dos projectes associats a un programa de màster o d'inici de doctorat. Per tant, existirà un procés de selecció de diferents candidats o candidates que siguin competents i mostrin un grau alt de motivació. La selecció de candidats o candidates incorporarà la perspectiva de gènere.

L'equip de treball format pel coordinador i els candidats(es) seleccionats(es) establirà una col·laboració continuada amb les entitats relacionades. Principalment, amb l'equip de govern encarregat de desenvolupar la plataforma online, incloent el grup Tecnopolítica IN3/UOC, Internet Interdisciplinary Institute (IN3) i l'associació aLabs, així com altres entitats relacionades. Aquesta col·laboració comportarà l'assistència a reunions periòdiques i l'elaboració d'informes on es faci explícit el progrés del projecte. En tot moment, es garantirà la privacitat i seguretat de les dades tractades.

La Taula 1 mostra una estimació de la cronologia del projecte en el cas que les dues tasques principals es desenvolupin en paral·lel per dues persones, juntament amb el coordinador.

Table 1: Cronologia aproximada del projecte

| Part del projecte | Accions relacionades | | | | | | |
|----------------------|----------------------|-----|---|---|---|-----|-----|
| Part I (secció 3.1) | 1.1 | 1.2 | | | | 1.3 | |
| Part II (secció 3.2) | 2.1 | 2.2 | | | | 2.3 | |
| Temps (mesos) | 1 | 2 | 3 | 4 | 5 | 6 | 7 → |

En cas que només un candidat o candidata sigui escollit, les tasques es redistribuiran seqüencialment i l'experiència adquirida durant la primera tasca permetrà reduir el temps de desenvolupament de la segona. El cas en el temps

total de desenvolupament mai sobrepassarà els 12 mesos. Al final de cada acció, es lliurarà un informe per fer explícit el seguiment i l'avaluació continuada del projecte. Al final de cada part, s'elaborarà un entregable que inclogui tots els resultats corresponents a la part desenvolupada.

Es preveu la publicació dels resultats científics en conferències i/o revistes científiques relacionades. Exemples de conferències poden ser *The Internet, Policy & Politics Conference*, *Knowledge Discovery and Data Mining* o *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery*.

6 Pressupost

El pressupost del projecte inclou principalment dues ajudes a estudiants que estiguin finalitzant els seus estudis de màster o començant el doctorat, així com despeses d'equipament informàtic i altres despeses associades a la publicació d'articles, com registres a conferències internacionals i viatges. La Taula 2 resumeix els costos.

Table 2: Pressupost

| Concepte | Cost |
|---|----------|
| Estudiant nivell master / inici PhD | 5,500 € |
| Estudiant nivell master / inici PhD | 5,500 € |
| Equipaments: dos laptops | 3,000 € |
| Despeses en registres conferències internacionals | 4,000 € |
| Total | 18,000 € |

7 Capacitats per l'elaboració del projecte

El sol·licitant del projecte té àmplia experiència en el tractament de grans volums de dades i una trajectòria científica internacional rellevant en l'àrea d'aprenentatge automàtic i de mètodes estadístics i computacionals. El projecte és desenvoluparà principalment en un entorn de treball universitari, on es disposen dels recursos necessaris per poder definir l'equip de treball de manera òptima i desenvolupar el projecte satisfactòriament.

Per raons d'anonimitat, aquesta secció està més elaborada en el document addicional, on figuren les dades de l'aplicant.

References

- [1] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.
- [2] C. Cortes, M. Mohri, and A. Muñoz Medina. Adaptation algorithm and theory based on generalized discrepancy. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 169–178. ACM, 2015.
- [3] C. Cortes, M. Mohri, M. Riley, and A. Rostamizadeh. Sample selection bias correction theory. In *International Conference on Algorithmic Learning Theory*, pages 38–53. Springer, 2008.
- [4] J. Huang, A. Gretton, K. M. Borgwardt, B. Schölkopf, and A. J. Smola. Correcting sample selection bias by unlabeled data. In *Advances in neural information processing systems*, pages 601–608, 2006.
- [5] Mobile World Capital, Barcelona. L’escletxa digital a Barcelona. <http://www.mobileworldcapital.com/escletxa-digital/>, 2016.
- [6] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [7] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000.