

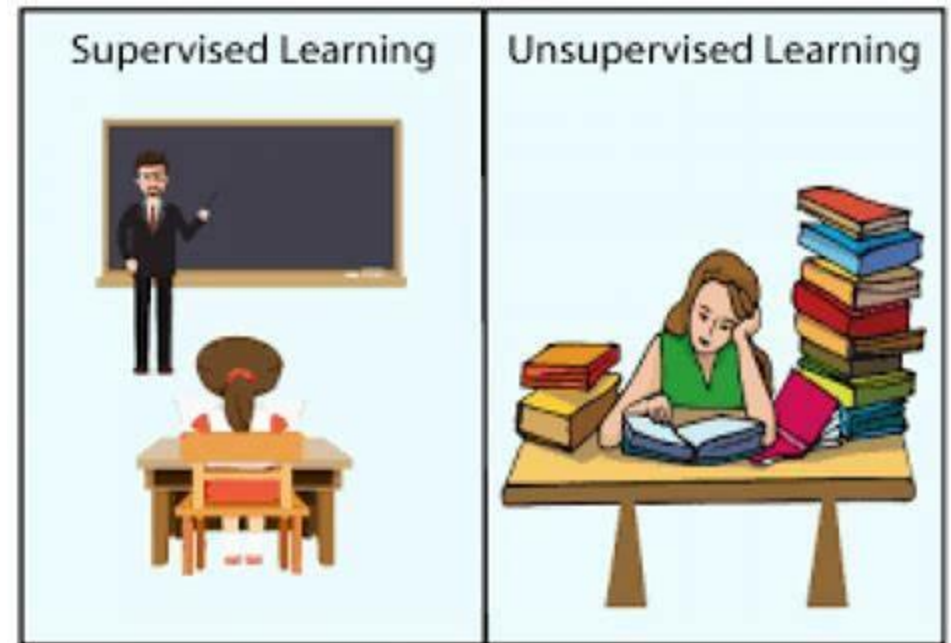
Dirbtinis intelektas

INF4042

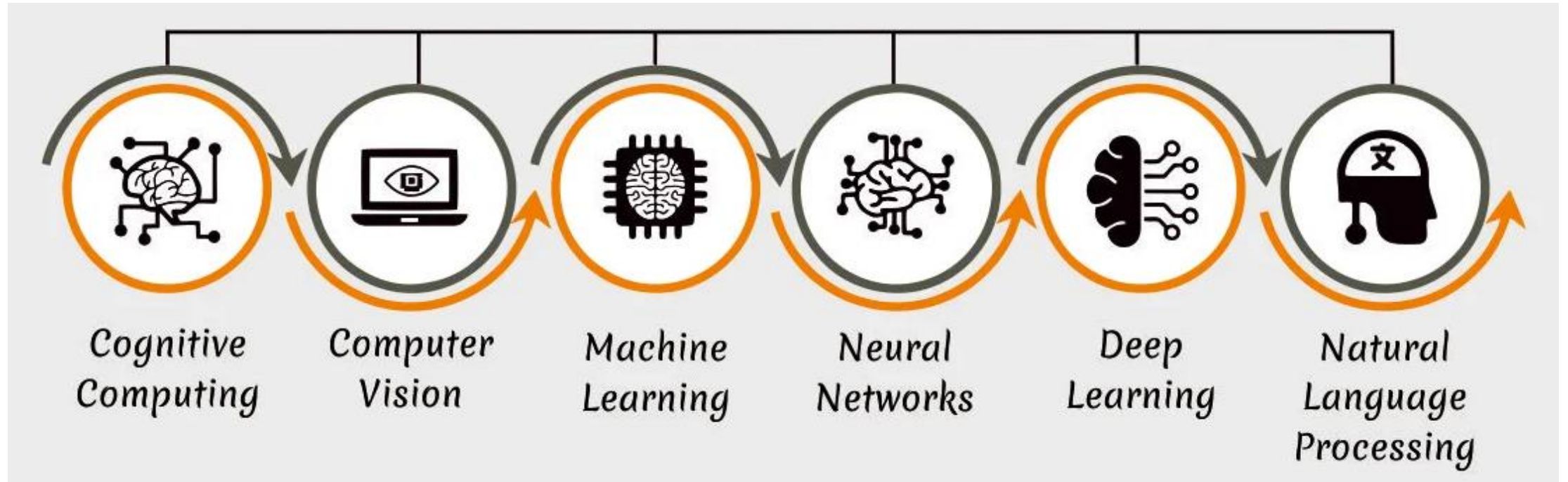
Parengė: dr. Aušra Vidugirienė

DI algoritmų tipai

1. Mokymas su mokytoju
2. Mokymas be mokytojo
3. Mokymas ir su mokytoju ir be mokytojo
4. Mokymas su pastiprinimu



DI algoritmai naudojami:



<https://www.pantechlearning.com/artificial-intelligence-algorithm/>

DI algoritmų tipai

- Mokymas su mokytoju:
 - Algoritmai su aiškiai sužymėtais apmokymo duomenimis, leidžiant programai mokytis ir pavyzdžių.
- Mokymas be mokytojo:
 - Algoritmai su nežymėtais apmokymo duomenimis, verčiant mokytis, nepriklausomai nuo sužymėjimo.
- Mokymas su pastiprinimu:
 - Algoritmai, kurie mokosi gaudami atgalinį ryšį iš ankstesnių veiksmų.

Regresija, klasifikavimas, mašininis mokymas

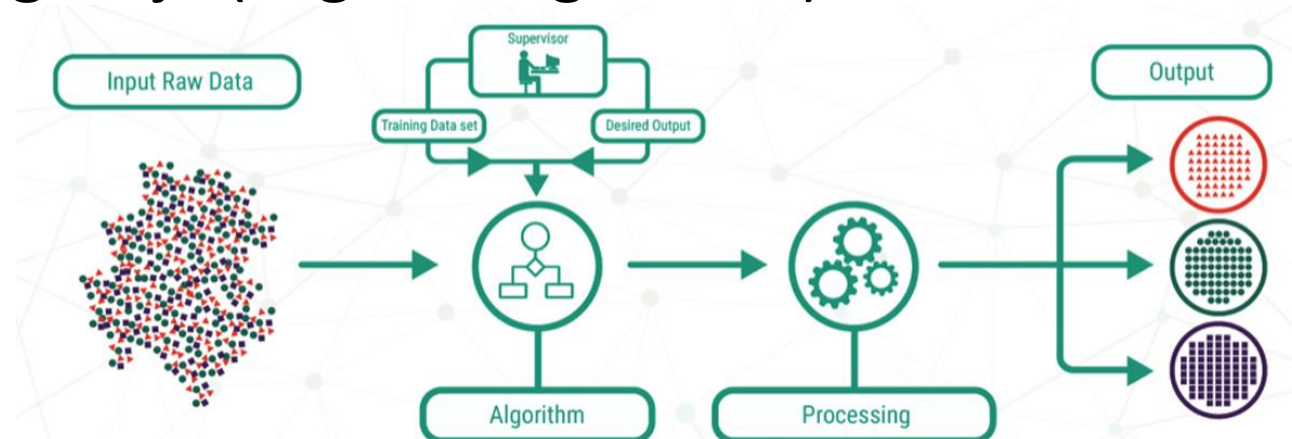
- **Regresija** yra koncepcija kai modeliuojamas santykis tarp išėjimo kintamojo (pvz.: automobilio kaina) ir vieno ar daugiau nepriklausomų kintamųjų (pvz.: markė, metai, spalva, kt.)
- **Klasifikavimas** yra koncepcija, kai kintamieji klasifikuojami į vieną ar daugiau rūšių, priklausomai nuo jų savybių (pvz.: markė: Toyota, Ford, Mazda; metai: 2000-2004, 2005-2010, 2011-2015)
- **Mašininis mokymas** yra koncepcija, kai kompiuterio programa mokoma mokytis ir adaptuotis regresijos ar klasifikavimo uždaviniuose be žmogaus įsikišimo.

Mokymo su mokytoju algoritmai

- Jie naudoja sužymėtus duomenis kitų duomenų rezultatų prognozei
- **Klasifikavimas** reiškia ir/ar rezultatą, naudojant dvejetainę sistemą (0 = no, 1 = taip). Taip pat egzistuoja **daugiaklasis klasifikavimas**, kur sprendžiamas duomenų skirstymas į nustatytas kategorijas ar tipus, priklausomai nuo uždavinio.
- **Regresija** reiškia, kad rezultatas bus realus skaičius (suapvalintas arba su skaičiais po kablelio). Paprastai naudojamas priklausomas kintamasis ir nepriklausomas kintamasis, o algoritmas panaudos abu variantus, kad apskaičiuotų rezultatą (prognozę arba apibendrintą įvertį).

Mokymo su mokytoju algoritmai

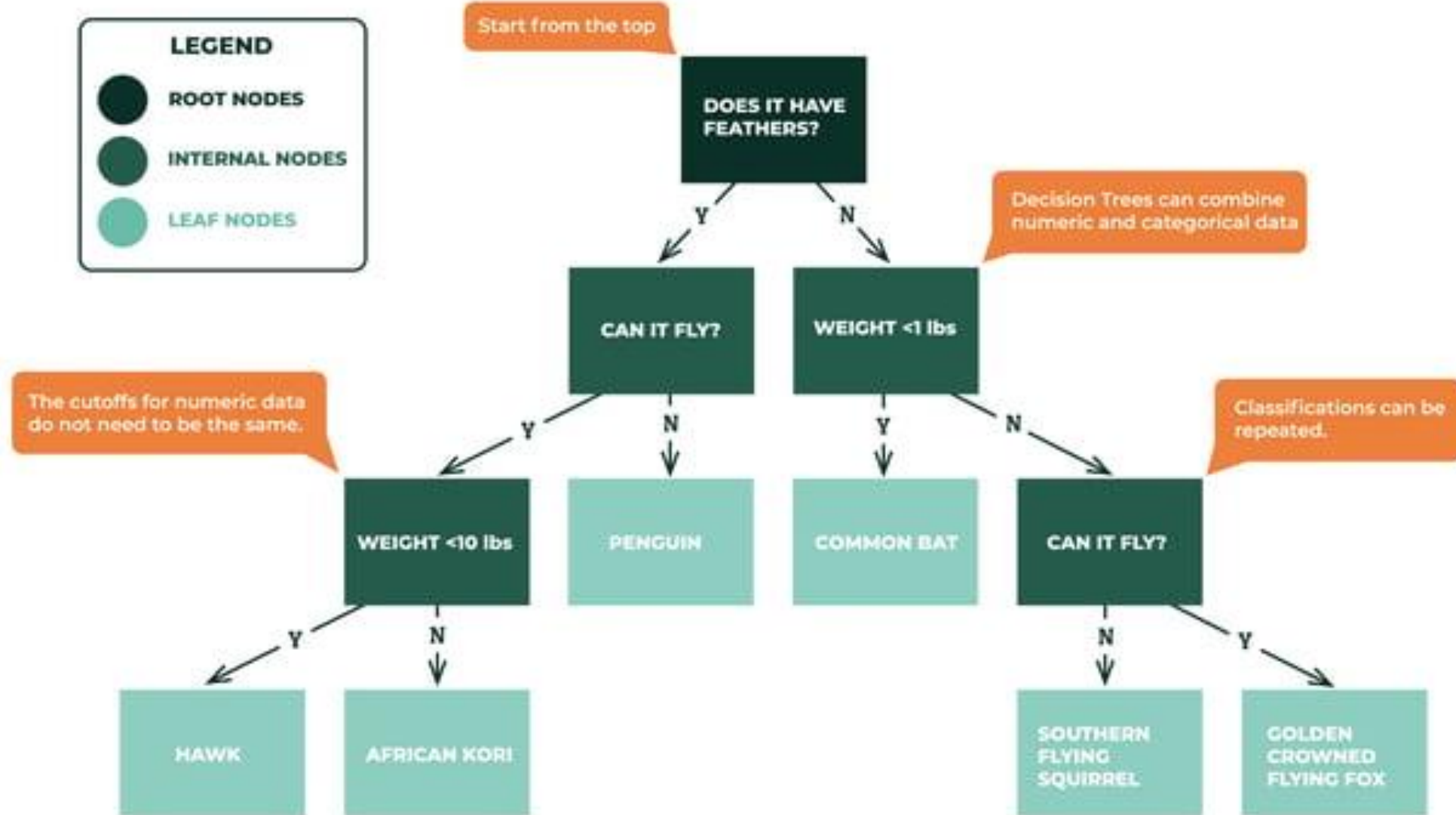
- Sprendimų medžiai (Decision tree)
- Atsitiktiniai miškai (Random forests)
- Atraminių vektorių mašinos (Support vector machines)
- Naivus Bajeso metodas (Naïve Bayes)
- Tiesinė regresija (Linear regression)
- Logistinė regresija (Logistic regression)



Sprendimų medžiai

- Sprendimų medžiai viską klasifikuoja į sprendimų mazgus.
- Jie naudoja selekcijos kriterijus vadinamus „atributų selekcijos matai“ (ASM) kurie atsižvelgia į įvairius matus (pvz. entropija, pelno koeficientas, informacijos nauda, kt.).
- Naudojant bazinius duomenis ir sekant ASM, sprendimų medis gali klasifikuoti gautus duomenis, dalindamas mokymo duomenis į sub-mazgus kol pasiekiamas išvada.

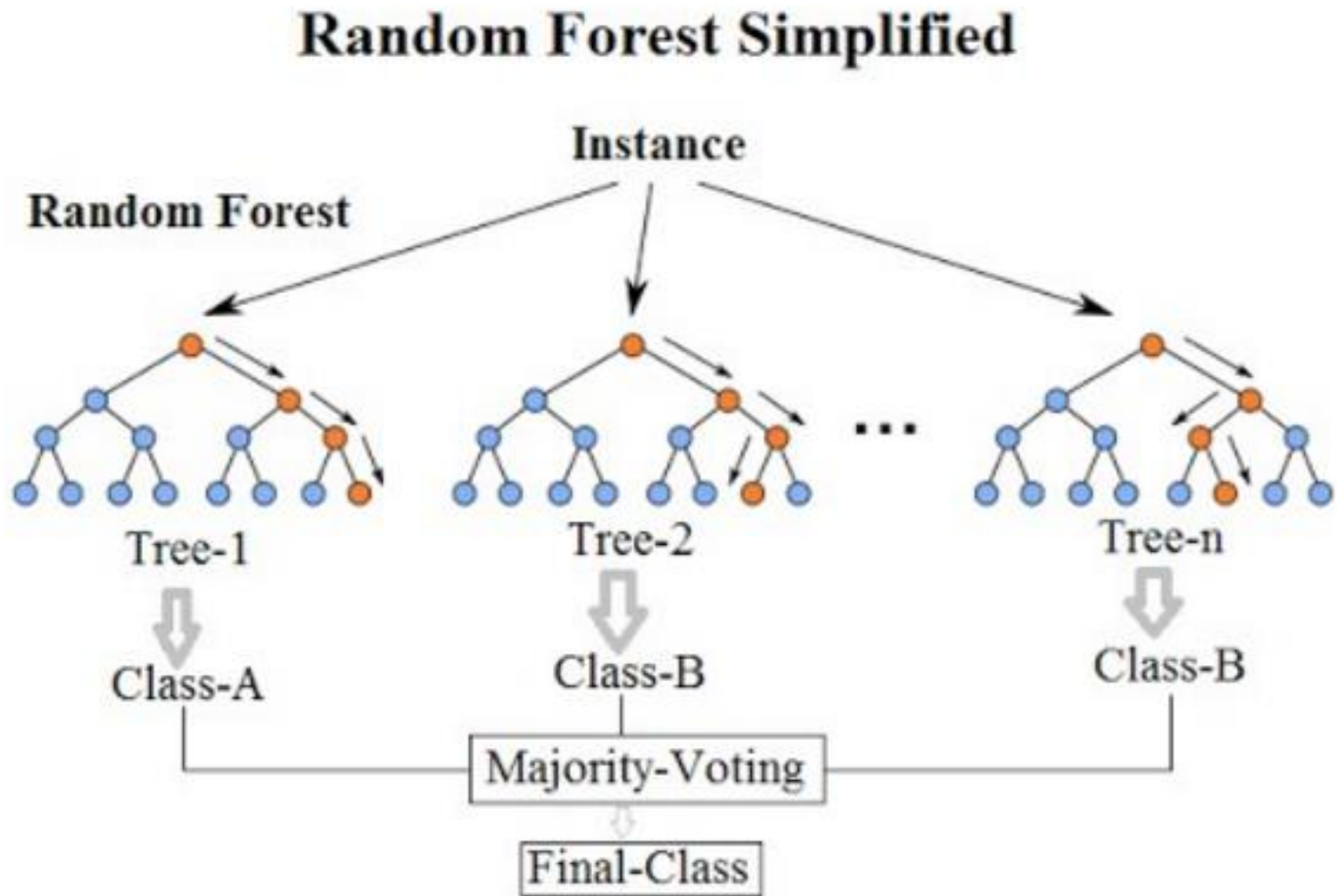
Sprendimų medis: pavyzdys



Atsitiktiniai miškai

- Atsitiktinių miškų algoritmas yra platus skirtingų sprendimų medžių algoritmų rinkinys.
- Atsitiktiniai miškai sukuria skirtingus sprendimų medžius, kad gautų dar tikslesnius rezultatus.
- Jie gali būti naudojami ir klasifikavimui, ir regresijai.

Atsitiktinis miškas: pavyzdys

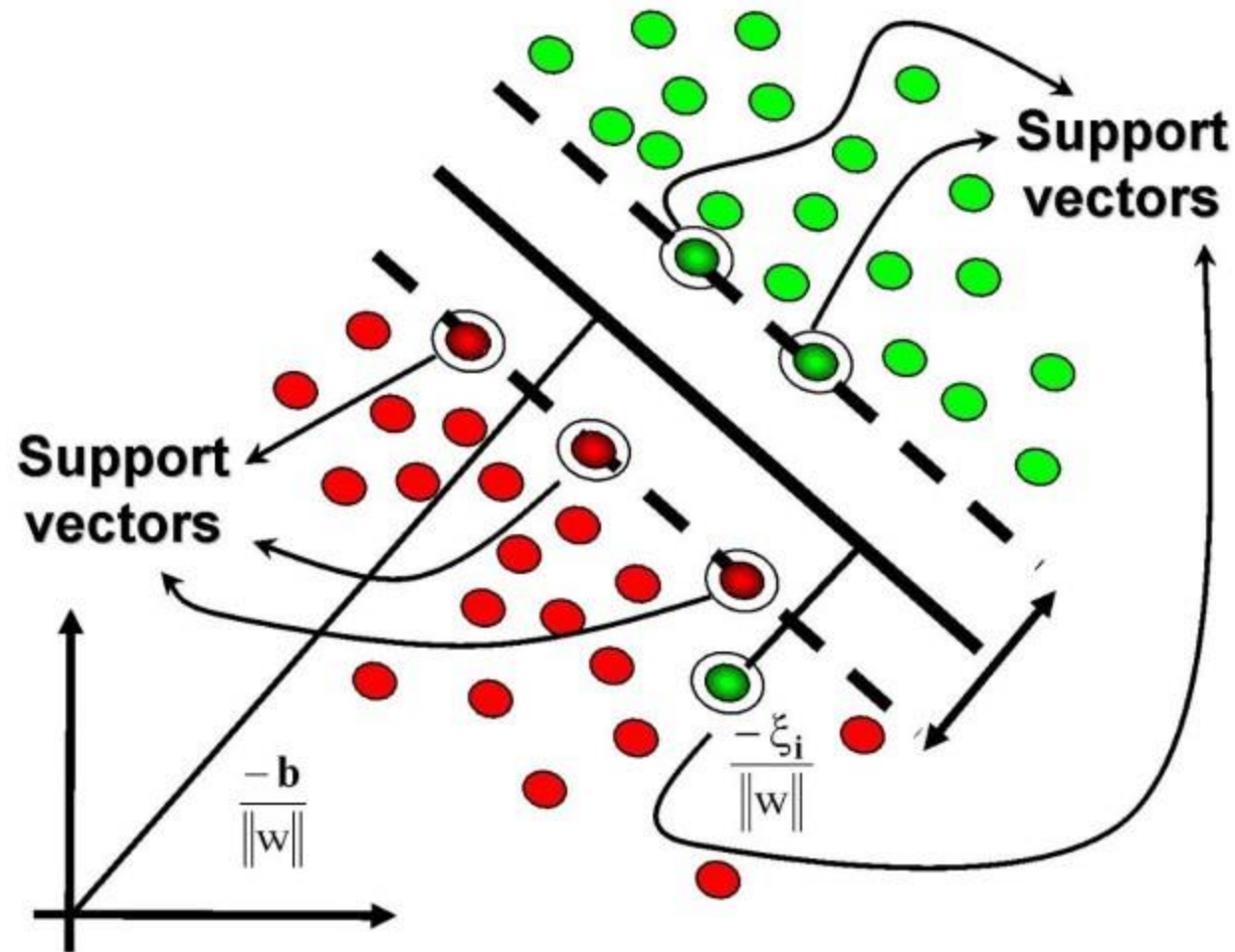


- <https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d>

Atraminių vektorių mašinos (angl. SMV)

- Atraminių vektorių mašina (angl. SVM) gali būti naudojama tiek klasifikavimui, tiek regresijai (nors dažniausiai naudojama klasifikavimui).
- Atraminių vektorių mašina veikia, patalpindama kiekvieną duomenų elementą į erdvę (n -matėje erdvėje, kur N – duomenų elementų skaičius).
- Tuomet algortimas klasifikuoja duomenų taškus, surasdamas hiperplokštumą, kuri juos atskiria. Gali būti daugiau nei viena hiperplokštuma.
- Atraminių vektorių mašina yra *artimiausio kaimyno* ir *tiesinės regresijos* metodų kombinacija.

Atraminių vektorių mašina (SMV): pavyzdys



- <https://educationalresearchtechniques.com/2016/07/15/basics-of-support-vector-machines/>

Naivusis Bajesas

- Priežastis, kodėl šis algoritmas buvo taip pavadintas yra ta, kad jis paremtas Bajeso teorema, o tuo pačiu pasikliauja prielaida, kad tarp požymių, esančių vienoje klasėje nėra ryšio.
- Šis algoritmas naudojamas didelėms duomenų imtims su dideliu kiekiu klasių. Tai – klasifikavimo algoritmas.

Naiviojo Bajeso klasifikatorius

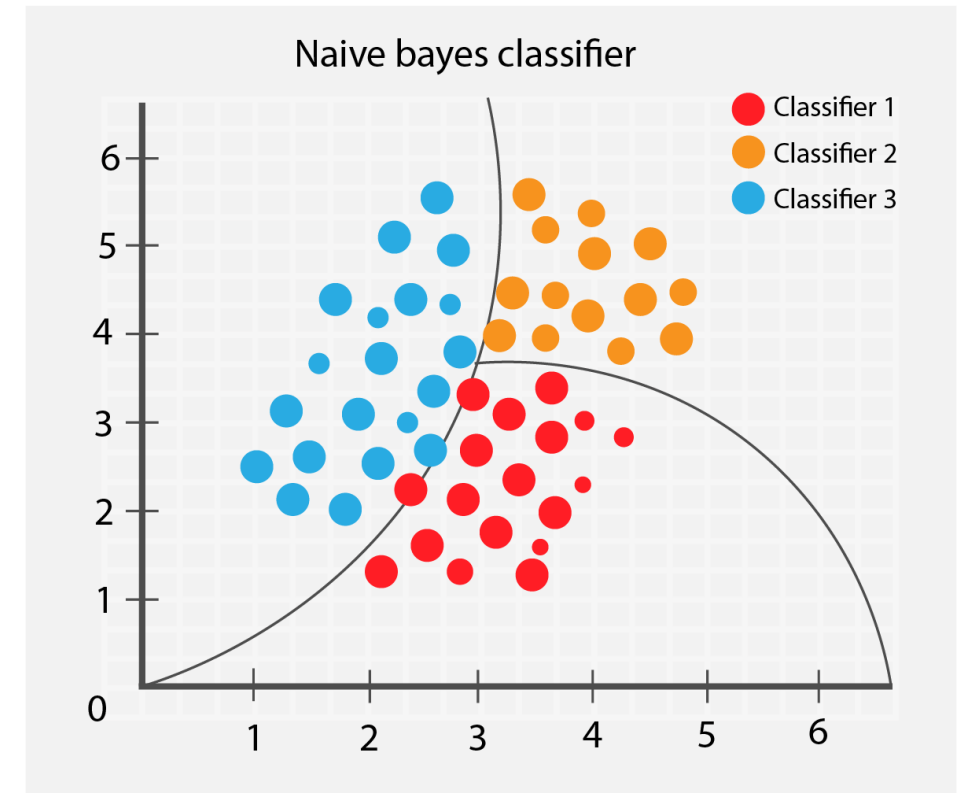
In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

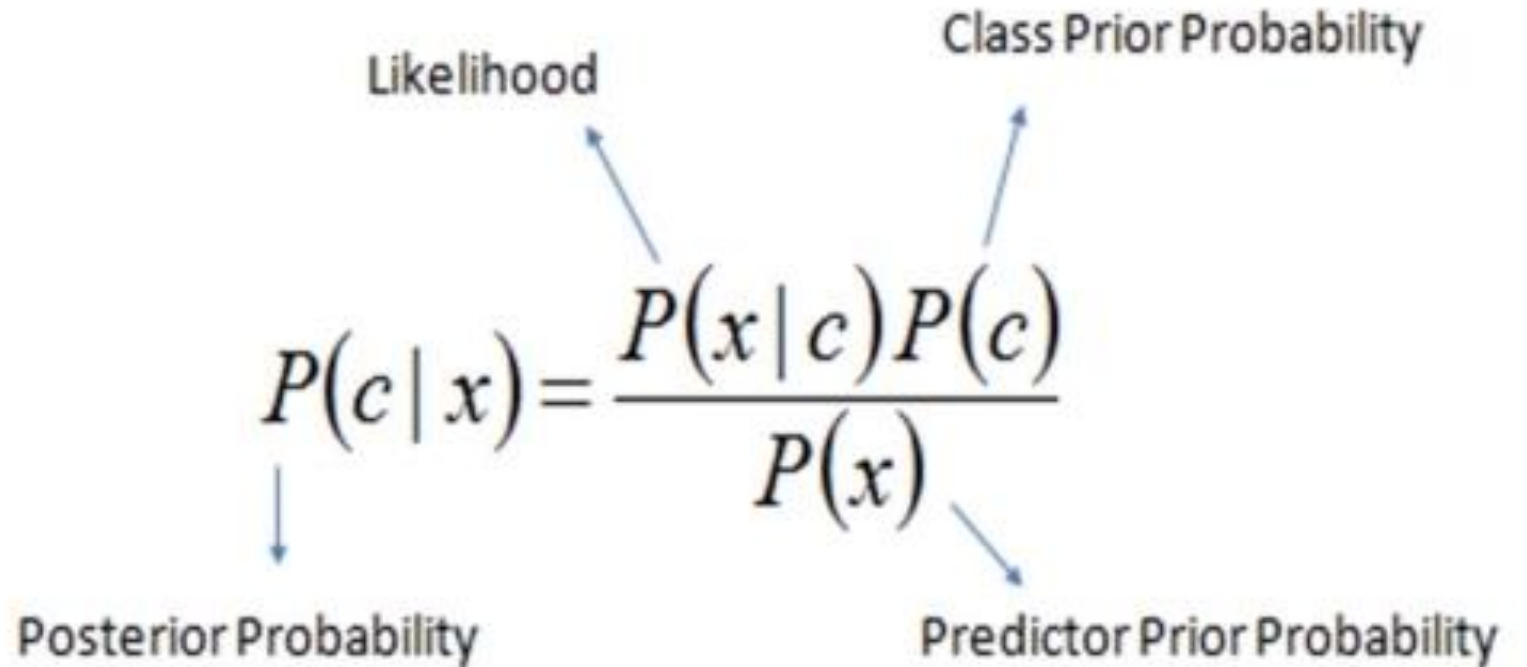
using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

- thatware.co



Naiviojo Bajeso klasifikatorius



The diagram shows the Bayes' theorem formula with four labels and arrows pointing to the corresponding parts of the equation:

- Likelihood** points to $P(x|c)$
- Class Prior Probability** points to $P(c)$
- Posterior Probability** points to $P(c|x)$
- Predictor Prior Probability** points to $P(x)$

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

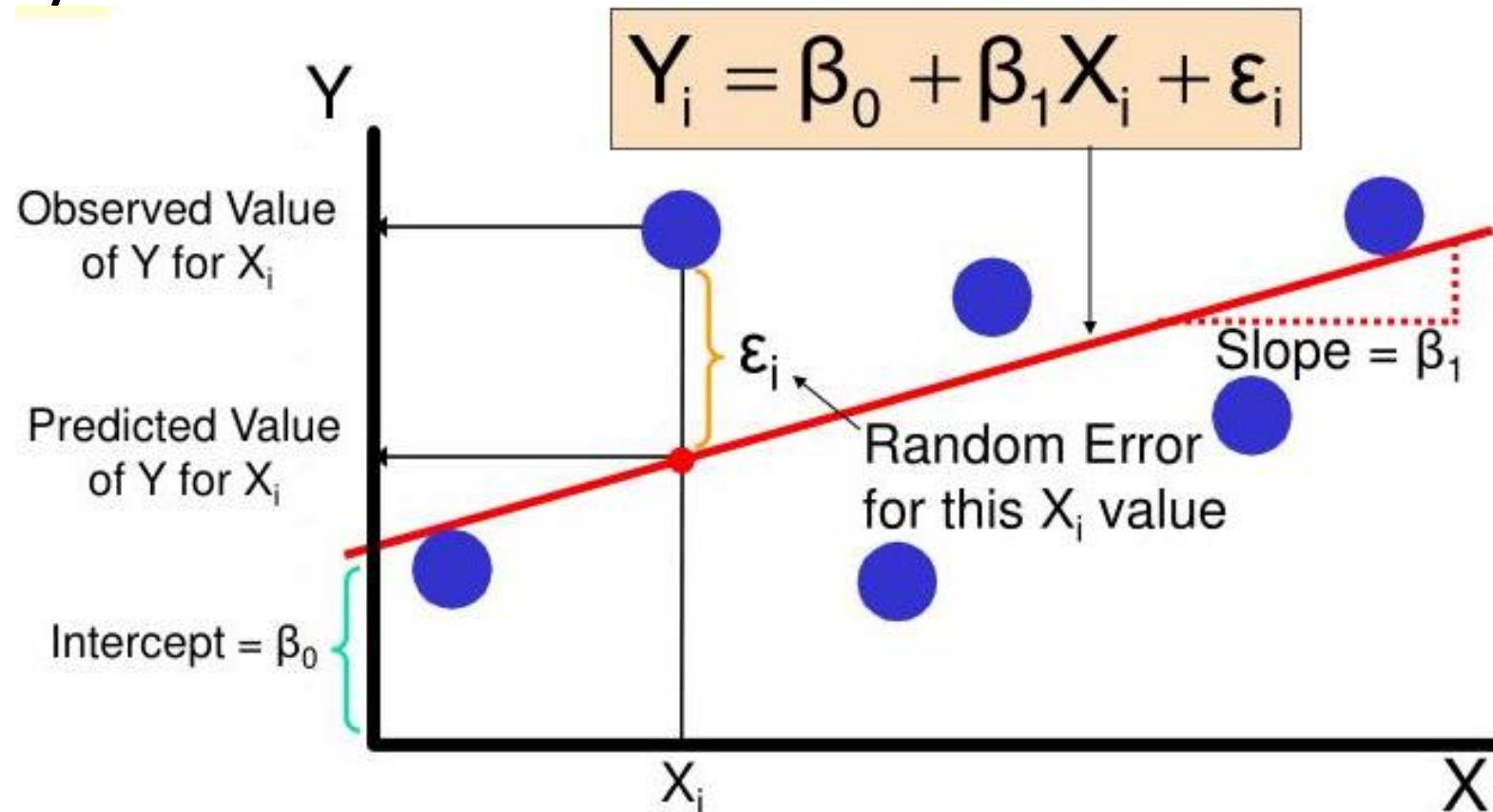
$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

- www.codespeedy.com

Tiesinė regresija

- Tiesinė regresija naudojama regresijos modeliavimui.
- Ji daugiausiai naudojama atrasti sąryšius tarp duomenų taškų ir prognozių.
- Panašiai kaip AVM, jis piešia duomenis plokštumoje, kur x ašyje – nepriklausomas kintamasis, o y ašyje – priklausomas kintamasis.
- Duomenų taškai atvaizduojami tiesiškai, kad būtų surasti jų sąryšiai ir galima būtų prognozuoti galimus tolesnius duomenis.

Paprastas tiesinės regresijos modelis: pavyzdys



- Daugiau informacijos: <https://www.slideserve.com/griffin-watson/chapter-4-5-24-simple-linear-regression>

Logistinė regresija

- Logistinės regresijos algoritmai paprastai naudoja binarę reikšmę (0/1) nepriklausomų kintamųjų aibės reikšmes įvertinti.
- Logistinės regresijos rezultatas yra 1 arba 0, taip ar ne.
- Kaip pavyzdys gali būti šlamšto filtras el.pašte. Pažymimas šlamštas (0) arba ne (1).
- Šis metodas naudingas tik tuomet, kai priklausomas kintamasis yra kategorinis (taip ar ne).

Tiesinė regresija prieš logistinę regresiją

