**Planning maken**

Voor elk model:

- trainen op trainset (geen validation set)

- probabilities op validation set berekenen 🡪 voor trainen van ensemble.

- probabilities op test set berekenen 🡪 verwerken in ensemble submissie.

- classificaties op test set berekenen 🡪 individuele submissie.

Afspraak: voeg ook de business ids toe aan de probabilities!

Validatie set/train set + businesses zonder labels!

Vrijdag: individuele modellen af

Maandag: finetunen van ensemble.

**Vragen Maaike**

**Gaussian mixture model: hoort dit zoveel geheugen in te nemen?**

Maar twee modellen op 64GB geheugen. En bovendien supertraag.

Is het normaal?

Kan best, aangezien alles met elkaar vermenigvuldigt. Dus dan explodeert het.

Opencv library. Ken je die? Beter?

Kan. sklearn python kant. Opencv, matlab. Durf niets over te zeggen.

Weet je zeker dat je de goede dingen erin stuurt?

Ja standaard functie. Er is niet veel wat er fout aan te doen is. Ander cluster algoritme opzoeken?

Maaike zal op github kijken, of ze er iets uit kan halen. Als het een standaard functie is, dan kan ze er niet veel aan doen.

Andere cluster soft assignment methode opzoeken.

misschien te veel dimensies? 4096.

Ja kan, alle datapunten keer alle datapunten is natuurlijk erg veel.

PCA? Kan.

Hij moet ook sneller, niet alleen geheugen.

Stuur welk script het is.

**De validatie businesses neem je dus niet mee bij het trainen vd modellen? Ook niet voor de uiteindelijke submissie?**

Dan gebruik je toch weer de hele set.

Dan kan de validatie set ook groter.

Geen overfitting, want ensemble traint niet.

**Tips voor het finetunen van de ensemble: langs alle mogelijke gewichten lopen en de score berekenen per combinatie. Lineaire regressie toepassen? Maar hoe kan dat 3D?**

Voor elk model een gewicht vinden. Kijken, als dit gewicht omhoog gaat, wordt het dan beter.

Majority voting 🡪 simpel

Regressie 🡪 ieder model eigen gewicht.

Maar dan moet je alle waardes langsgaan.

Gradient descent. Regressie.

Maar pas op, niet overfitten op de validatieset.

**Modellen:**

Color feature + random forest

- trainen op alleen trainset

- probabilities op validation set berekenen

- probabilities op validation set berekenen

- classificaties op test set berekenen 🡪 individuele submissie.

Caffe feature modellen

SVM 1.0 + mean:[Danielle]

- niet nodig indien SVM 2.0 werkt.

SVM 2.0 + Q123: [Diede]

- trainen op alleen trainset

- probabilities op validation set berekenen

- probabilities op validation set berekenen

- classificaties op test set berekenen 🡪 individuele submissie.

Cluster model: [Laurens]

- k-means clusters. Aantal clusters was 100. Maar niet succesvol.

- Gaussian mixture model werkt niet. Iets anders proberen.

SVM 3.0 + per foto trainen: [Daniëlle][Roos]

- trainen op alleen trainset

- per foto classificatie omzetten naar business classificatie

- probabilities op validation set berekenen

- probabilities op validation set berekenen

- classificaties op test set berekenen 🡪 individuele submissie.

Ensemble: [Roos]