

## 1. Doelstelling:

De eerste twee weken hebben wij besteed aan de data bestuderen en een hoofddoel formuleren. We wilden de data inkrimpen, dus besloten alleen naar avondeten te kijken. Het hoofddoel die we toen geformuleerd hebben was:

- Hoe kunnen we een app maken waarmee gebruikers een persoonlijk advies krijgen wat ze deze week kunnen eten en die zich kan aanpassen op voorkeuren van gebruikers?
- Op basis van favorieten van gebruiker, avondmaaltijden voor een week suggereren kijkend naar de bereidingstijd.

Gedurende het project is dit nog veranderd en we hebben nooit meer iets met de bereidingstijd gedaan, omdat we niet zo ver kwamen en ons meer gefocust hebben op de keukens voorspellen. De doelstelling voor de tweede periode van het project was:

- Kunnen we op basis van de ingrediënten die in een recept zit voorspellen tot welke keuken een recept behoort en zo recepten aanraden aan gebruikers met keukens als tags ?

## 2. Data Preprocessing

De eerste periode bestond voornamelijk uit de data bestuderen en bewerken. Er waren 4 verschillende datasets: Tags (met alle recepten en bijbehorende tags, zoals "italiaans", "vegetarisch" etc.), Recipes (met plaatjes en linkjes naar de recepten en een paar gegevens, zoals voor hoeveel personen en hoeveel calorieën), Nutritions (de nutritions per recept, dus hoeveel verzadigd vet, hoeveel eiwit, etc.), Ingredients (met van elk recept alle ingrediënten en hoeveel daarvan)

Voorbeelden van wat ik met de data heb gedaan in de ontdekkingsfase (In het 3<sup>e</sup> hoofdstuk 'Notebooks' zijn daar sommige voorbeelden van te vinden):

- Alle datasets ingeladen en bestudeerd. Gekeken naar de grootte en de gegevens.
- Nutritions gepivot en gegevens numeriek gemaakt om er visualisaties bij te kunnen maken en een simpel Lineair Regressie model te leren fitten op een echte dataset.

*Met de groep besloten we aan de slag te gaan met de datasets Tags & Ingredients. Als we voorspellingen (Classification models) wilde maken over recepten op basis van welke tags bij een recept hoorde of welke ingrediënten erin zaten, moest de data eerst bewerkt worden (Dummy data). Het moest numeriek gemaakt worden.*

- Ik heb Tags omgezet naar een Dataframe met de recepten als index in de rijen en de tags als kolommen. Vervolgens liet ik in de dataframe een 1 invullen als een tag bij een recept voorkwam en een 0 als dit niet het geval was. Op deze manier is de data bruikbaar voor een model. Hetzelfde heb ik voor Ingredients gedaan.

*We ontdekten dat er in de dataset Ingredients veel ingrediënten voorkwamen die sterk op elkaar leken. Soms door haakjes, meervoud of verkleinwoordjes en soms doordat er meerdere soorten van iets zijn die heel erg op elkaar lijken. We wilden toen de dataset zo bewerken dat ingrediënten die sterk op elkaar leken, gecategoriseerd waren tot 1 ingrediënt.*

- d. Ik heb in een notebook laten zien, hoeveel ingrediënten sterk op elkaar leken. Met de library " " heb ik een functie geschreven die ingredienten die sterk op elkaar leken samen voegde. Micheal was hier ook mee bezig en we kwamen er samen achter dat er af en toe ingrediënten waren die onterecht samengevoegd werden, zoals "spruitjes" en "uitjes". Daarom hebben we de functie zo uitgebreid dat de dataset naar het engels vertaald werd en vervolgens nog een keer het proces werd uitgevoerd. Alleen ingrediënten die in beide talen samengevoegd werden, werden ook echt aangepast

*Uiteindelijk is dit nooit gebruikt, aangezien we na aanleiding van de interne presentatie en een gesprek met Jeroen besloten dat dit niks zou toevoegen aan ons model. Jeroen sprak uit ervaring en legde uit dat het vaak niet voor verbeteringen zorgde en dat er ook recepten waren waarvoor het juist nadelig was om die kleine verschillen eruit te halen.*

Voor dit project wilden we gaan werken met gesimuleerde gebruikers. We wilde de gebruikers in eerste instantie 10 favorieten meegeven en dan voorspellen met classificatie of die een recept wel of niet lekker zou vinden. Daarnaast wilde we in de train en testset zorgen dat er recepten zaten die een gebruiker lekker vond en juist niet, zodat die 0en (niet lekker) en 1en (wel lekker) kon voorspellen tijdens het trainen en testen.

- e. We hebben het lang gehad over hoe de input data er nou uit zou zien en dat was moeilijk helder te krijgen. Ik heb toen een visualistie gemaakt van hoe de trainset er ongeveer uit zou zien, die we ook tijdens presentaties gebruikt hebben:

		ui	kipfilet	peper	...	ui	kipfilet	peper	...	y	random	One-out
1	user1				...				...	1	NaN	Rec1
2	User1				...				...	0	rndm1	Rec1
3	User1				...				...	1	NaN	Rec2
4	User1				...				...	0	rndm2	Rec2
5	User1				...				...	1	NaN	Rec3
6	User1				...				...	0	rndm3	Rec3
...	....	...	...	...	...	...	...	...	...	...	...	...

Tags van 7 favorieten
Van L.O.O. of Random recept

Ik heb, net als andere groepsgenoten, gepuzzeld met het schrijven van functies voor gesimuleerde gebruikers en voor het maken van zo'n dataframe als hierboven. Uiteindelijk had Bonno als eerst een snelle handige manier om dit te maken en we hebben toen de rest van het project zijn functies en gesimuleerde gebruikers gebruikt.

*Op een gegeven moment hadden we bedacht dat we dus naar de ingredienten kijken om te voorspellen welke keuken een recept is, zo kan het model herkennen of een recept bij een bepaalde smaak van een gebruiker past. Stel de gebruiker houdt van italiaanse recepten, dan zal het model snel een recept aanraden waarvan het denkt dat het uit dezelfde keuken komt.*

- f. Er zijn nog meer verschillende ingrediënten dan tags, dus leek het ons handig om wat te krimpen en niet alle ingredienten mee te nemen in het model. Ik heb de dataset ingekort en alleen de meest gebruikte

### 3. Predictive Analyses

Als de data klaargemaakt was om te gebruiken in classificatie modellen, konden we werken aan het voorspellingsmodel. We begonnen simpel en los van elkaar, maar hebben de laatste paar (4) weken gezamenlijk gewerkt aan het eindmodel. Bonno heeft grotendeels de code daarvoor geschreven, op een gegeven moment vond ik het moeilijk veel te helpen, omdat er steeds opeens superveel code bijkwam en ik dat niet zo snel kon begrijpen. Daarom heb ik soms meer m'n eigen ding gedaan en het wat simpeler gehouden. Daarnaast heb ik tijdens de meetings wel goed nagedacht over wat handig was en de resultaten van modellen geanalyseerd. Ik vind zelf dat ik met nuttige toevoegingen kwam tijdens de meeting met docenten en zo veel heb geholpen aan het verbeteren van het model. Ook heb ik tussendoor kleine functies geschreven die anderen toegevoegd hebben aan het grote eindmodel.

Voorbeelden van wat ik gedaan heb tijdens het maken van voorspellingen en werken aan ons model (In het 3<sup>e</sup> hoofdstuk 'Notebooks' zijn daar sommige voorbeelden van te vinden):

- a. Simpele Lineaire Regressie – df Nutritions.  
Om te oefenen met het trainen van een model op een echte dataset, heb ik gekeken naar koolhydraten en energie (in kcal) in recepten. Ik heb een scatterplot gemaakt en vervolgens een simpel Lineair Regressie model gefit, die voorspeld hoeveel calorieën er in een recept zitten op basis van de hoeveelheid koolhydraten. Dit is gelukt, maar het model had wel een hele hoge RMSE dus is zeker nog niet optimaal.
- b. Tomaten classificatie model – df Ingredients  
Omdat wij in ons project werken met classificatiemodellen heb ik vervolgens ook een simpel model gemaakt wat van 1 ingredient voorspelt of het wel of niet in het recept zit op basis van andere ingrediënten. Ik heb drie verschillende modellen gefit (Logistic Regression, K-Nearest Neighbour en Decision Tree) en deze vergeleken. Ook hier zijn de resultaten niet zo goed, maar finetunen van de modellen kan helpen.
- c. Werken aan het eindmodel samen met de groep.

### 4. Resultaten

Het model had uiteindelijk een accuracy van 99.7 model.

Ook is er door Michael en Mohamed een applicatie gemaakt. In de applicatie kan een user een keuken opgeven die die lekker vindt en 10 recepten 'leken'. Zo kan de app, als het model gekoppeld is, vervolgens voorspellen welke gerechten de user lekker zou vinden en dit in een weekmenu gooien. Het weekmenu wordt zo opgebouwd dat er een enkele favoriet ook in zit, naast recepten die het model voorspelt. De user kan het weekmenu zien en alle details over de recepten bekijken.

### 5. Terugkoppeling doelstelling

Uiteindelijk is onze doelstelling gedurende het project dus veranderd. Het doel wat wij voor de eerste periode hadden is niet behaald. Het model wat wij wilde ontwikkelen dat op basis van favorieten van gesimuleerde gebruikers recepten aanraadt kreeg niet een goed accuracy score en na veel finetuning en het proberen van veel verschillende modellen besloten we te beginnen met een andere aanpak.

Het doel wat we in de tweede periode opgesteld hadden is wel behaald. Er is een model wat op basis van een paar keukens kan voorspellen of een user een recept lekker zal vinden.

De applicatie is een leuke visualisatie om te laten zien wat het model kan, maar is niet helemaal af aangezien het niet gekoppeld is aan het model.

## 6. Aanbevelingen

Applicatie werkt en het model werkt, maar ze zijn nog niet gekoppeld aan elkaar.

Aanbeveling: Onderzoek doen naar hoe je de applicatie goed op het model aansluit.

Vervolgens is het een interessante stap om de applicatie te laten testen door test users.

Op dit moment wordt er nog niet rekening gehouden met andere dieetwensen van gebruikers, maar alleen naar keukens die zij wel/niet lekker vinden.

Aanbeveling: Meerdere dieetwensen mee laten geven en het model rekening laten houden met meerdere opties. In vervolg onderzoek kan worden onderzocht welke factoren belangrijk zijn om mee te nemen en hoe dat in het model verwerkt kan worden

Ook is het belangrijk dat mensen die de app gebruiken een doel kunnen opstellen en dit na kunnen streven en dat zij in een week gevarieerd eten, zodat ze hun dieet niet snel zat worden. Het enige wat nu mee genomen wordt in de app, is dat je elke week favorieten recepten en een paar nieuwe recepten eet.

Aanbeveling: Mensen een doel, zoals minder calorieën, laten instellen en onderzoek doen naar hoe dit meegenomen kan worden in het model. Dit is mogelijk aan te pakken met Lineair programmeren.

Aanbeveling: Onderzoek doen naar hoe je een gevarieerd weekmenu creëert en hoe dit in het model verwerkt kan worden. Zorg bijvoorbeeld dat een user niet 4 keer in de week pasta eet of iets met veel vet.