**Voorstel model ontwikkeling Bastian Lewis**

### Introductie

### In dit voorstel omschrijven we de taken en de verwachte tijd voor de verschillende taken. Deze schattingen zijn gemaakt op basis van de verstuurde data en het gesprek met Timo Archutowski. Deze schattingen zijn naar beste inzicht. Echter kan dit altijd langer of korter uit vallen. In dit geval zal contact worden opgenomen met de opdrachtgever om te bespreken hoe het project voort te zetten.

### In dit voorstel gaan we uit van de ontwikkeling van één model. Dit is een model om te voorspellen wat de kans is dat iemand starter wordt. Indien ook een model ontwikkelt dient te worden om de waarde van een dergelijke starter te schatten (berekening) dan dient hiervoor een nieuw model ontwikkeld te worden. Voor een kostenschatting hiervan kunnen de kosten schattingen voor modellering worden genomen.

### Het is belangrijk om te beseffen dat op voorhand geen uitspraken gedaan kunnen worden over de kwaliteit of bruikbaarheid van het te ontwikkelen model. Na de exploratieve analyse dient eerst een evaluatie plaats te vinden om te bepalen of het überhaupt zinvol is een model te ontwikkelen. Na de eerste modelering-ronde dient bepaald te worden of het ontwikkelde model voldoet aan de verwachtingen of dat er een nieuwe modelering-ronde dient te volgen of dat het project eventueel dient te worden stopgezet.

### Overzicht van taken

|  |  |  |
| --- | --- | --- |
| Taak | Geschatte uren | Output |
| Data cleaning | 16 | Opgeschoonde dataset |
| Exploratieve analyse | 8 | Rapportage descriptieve data en correlaties |
| EVALUATIE MOMENT 1 | | |
| Featue engineering | 8 | Feature dataset (for modelling purposes) |
| Modelering ronde 1 | 16 | Model\_0 + Rapportage model performance |
| EVALUATIE MOMENT 2 | | |
| Featue engineering | 4 | Feature dataset (for modelling purposes) |
| Modelering ronde 2 | 16 | Model\_1 + Rapportage model performance |

### Tijdens de verschillende evaluatie momenten kan door de opdrachtgever besloten worden het project voortijdig te beëindigen. Dit kan op basis van adviezen van de opdrachtnemer maar ook naar eigen inzicht van de opdrachtgever.

### *\*NOTE: Dit voorstel is bedoelt om snel een model te ontwikkelen dat mogelijk bruikbaar kan zijn in productie. Normaal gesproken duurt een dergelijk proces aanzienlijk langer en is uitgebreider. Het risico om fouten bij een snelle ontwikkeling van het model of latere problemen met toepassing van het model zijn groter.*

### Totale kosten

### Geschatte tijd voor opleveren van een eerste model (model\_0) is 48 uur. De kosten bij een uurtarief van 90 euro per uur zin dan in totaal €4320,- exclusief btw.

### Geschatte tijd voor opleveren van een tweede model (model\_1) indien gewenst is 60 uur. De kosten bij een uurtarief van 90 euro per uur zin dan in totaal €5400,- exclusief btw.

### *\*NOTE: alle schattingen zijn op basis van een positief scenario. Er moet rekening mee gehouden worden dat dit niet altijd het geval is. Bij een verwachte urenoverschrijding van meer dan 10% op enig moment in het proces zal de opdrachtnemer contact opnemen met de opdrachtgever om te bepalen hoe verder te gaan.*

### Alternatieve voorstellen

### Een dergelijk snelle ontwikkeling van een model kan zeer waardevol zijn om te bekijken of het überhaupt mogelijk en zinvol is met de huidige data een goed model te ontwikkelen. Echter het risico op fouten tijdens de data cleaning en modelontwikkeling zijn groter gezien de tijdsdruk wat later tot problemen kan leiden.

### Er kan ook worden gekozen om het proces uitgebreid te doorlopen door meer tijd te besteden aan alle verschillende stappen. Dit kan zowel direct gedaan worden maar ook achteraf om het model te optimaliseren.

### Op deze manier kunnen fouten worden voorkomen of gecorrigeerd worden en kan het meest optimale model gevonden worden. Voor een dergelijke uitgebreid proces zou ik zeker 4 weken uittrekken (€14400,-).

### Daarnaast is het mogelijk zinvol om te kijken of er andere of meer data kan worden gebruikt om het model te verbeteren. Bijvoorbeeld door een aanvullend invulformulier of het scrapen van externe databronnen zoals LinkedIn.