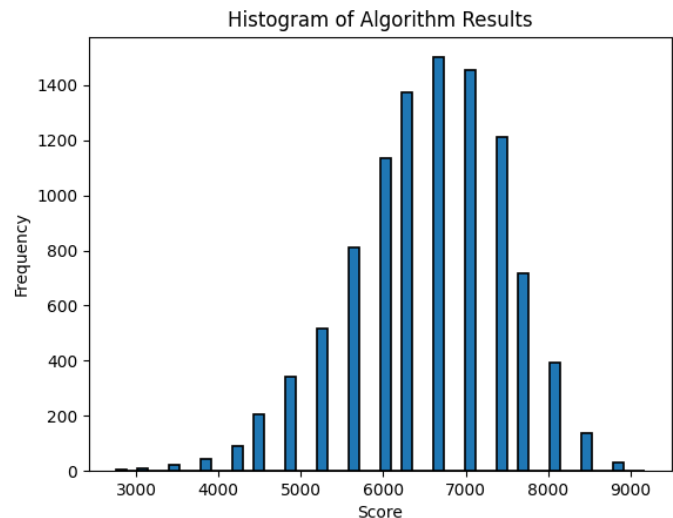


Baseline RailNL Dapper

Voor de baseline hebben we een random algoritme ontwikkeld. Deze heeft als input 7 trajecten, een tijdsframe van 120 minuten en rijdt over Noord- en Zuid Holland. Een traject begint in een willekeurige station en loopt naar willekeurige andere stations. Er zijn géén constraints. Daarvoor is gekozen omdat dit het slechtste algoritme is aangezien er geen aspecten zijn toegevoegd om de trajecten beter te laten verlopen.

De histogram in afbeelding 1 laat de verdeling zien van de waarschijnlijkheid dat een bepaalde railnetwerk kwaliteit score voorkomt wanneer het algoritme random wordt gebruikt.



Afb1: Histogram van 10.000 iteraties van het algoritme random.py

Het laat een normale verdeling zien. Dit houdt in dat er een hogere frequentie is van scores middenin het bereik dan aan de onderzijde of bovenzijde. Het is een random algoritme, wat in eerste instantie doet verwachten dat er een uniforme steekproef plaatsvindt met dusdanig veel iteraties dat dit zou leiden tot een uniforme verdeling. Dat is nu echter niet het geval.

Voorwaarden voor een discrete uniforme verdeling is dat de steekproef onderling onafhankelijk en aselekt is. (1) Onderling onafhankelijk houdt in dat steekproeven elkaar niet beïnvloeden. (2) Deze voorwaarde wordt aan voldaan, omdat elke random score berekening opnieuw wordt gerund. Het probleem ligt bij het aselekt zijn. Een aselechte steekproef houdt in dat elke steekproef evenveel kans heeft om gekozen te worden. (2) Dit is voor het beginstation natuurlijk wel zo, maar vanaf dit station wordt een bepaalde route gekozen en daarop wordt de score zoals in de histogram bepaald. Stations verschillen in connecties, het ene station heeft er meer dan het andere. Hierdoor is het aannemelijk dat er over meerdere steekproeven heen, er vaker over bepaalde stations gegaan wordt dan over andere. Hierdoor is een normaal verdeling een logische uitkomst.

Een ander opvallend punt is dat de histogram niet volledig aaneensluitend is, maar 'gaten' bevat. Dit is een gevolg van de berekening van de score, die het aantal bereden verbindingen neemt ten opzichte van het totaal aantal verbindingen. Er zijn maar een bepaalde hoeveelheid verschillende trajecten die random kan maken met de restricties van trajectenaantal en tijd, waardoor er vaker exact dezelfde score uit kan komen.

Bronvermelding:

1. <https://www.math.leidenuniv.nl/~spieksma/colleges/kans-inf/dictaat.pdf>
2. <https://www.scribbr.nl/onderzoeksmethoden/steekproef-in-je-scriptie/>