1. **Intrinsic Evaluation**

In this section tackles the evaluation of words on our trained word2vec models based on semantics. Using these sample data, it will test the model’s output on how it is closer to a human judgment. The similarity percentage tells how close the relevance of two words.

We have randomly tested the model by setting 20 random pairs of IT skills.

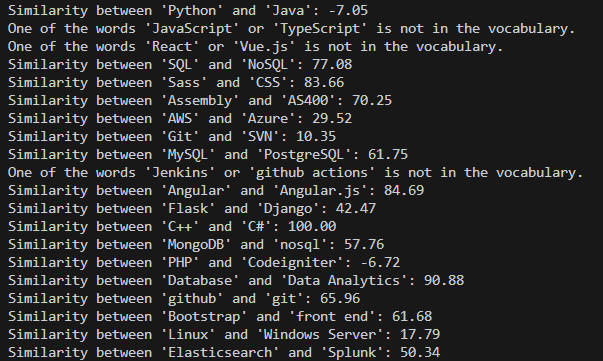


Figure A1: Result of similarity test percentage (IT related skills)

This shows that two relevant IT skill mostly scored higher on similarity test. Words with similar letters or almost matches the other one almost scored higher such as ‘C++’ and ‘C#’, ‘Database’ and ‘Data Analytics’, ‘Angular’ and ‘Angular.js’. Some skills are scored lower than 84% like ‘Sass’ and ‘CSS’ (83.66%), ‘Assembly’ and ‘AS400’ (70.25%), ’github’ and ‘git’ (65.95%), ‘Bootstrap’ and ‘Front end’ (61.98%),’MySQL’ and ‘PostgreSQL’ (61.75%), ‘MongoDB’ and ‘nosql’ (57.76%), ‘Elasticsearch’ and ‘Splunk’ (50.34%) ,‘Flask’ and ‘Django’ (42.47%), but higher than other pairs which is completely different such as ‘Linux’ and ‘Windows Server’ (17.79%), although Linux and Windows Server both kernel but they have different operation features which can dictate users paradigm or practice, a skill in Linux is not directly transferable to a skill in Windows Server and one factor is their CLI commands which is completely different that affect developer’s preferences and practices. Some results like words not found on the vocabulary and one pair which is obviously related which are ‘PHP’ and ‘CodeIgniter’ which is a framework based on PHP is set for recommendation for improvement and fine tuning of the model.

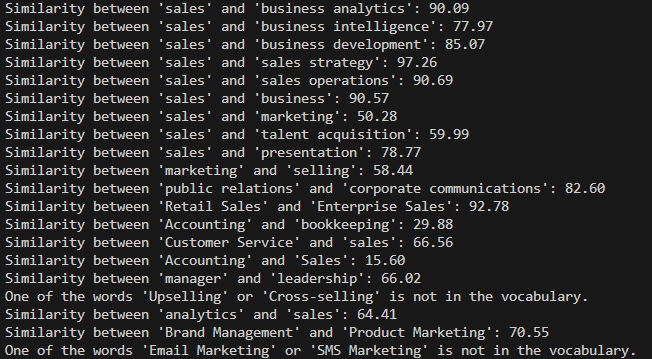


Figure A2: Result of similarity test percentage (Sales, Business and Marketing related skills)

This result shows the similarity percentage of various Sales, Business and Marketing related skills in words. We have tested 20 random pairs and concluded that based on observation, words which has similar letters or matches the other one scored higher in similarity test. It is also noticeably words with ‘sales’ and ‘business’ compared by pairs scored above 84% such as ‘sales’ and ‘business analytics’ (90.09%), ‘sales’ and ‘business development’ (85.07%) and ‘sales’ and ‘business intelligence’ (77.97%). One comparison with excellent similarity test percentage with no similar word matches score also higher than 80% which is the ‘public relations’ and ‘corporate communications’ where in a sense, ‘relations’ and ‘communications’ are relatively in a closer context. ‘Accounting’ and ‘Bookkeeping’ may seems similar to some people or in a general perspective but actually have hard line difference where Accounting is about analyzing, reporting and interpreting financial data while the bookkeeping is more on recording daily financial transactions, we can say that bookkeeping shares under the same umbrella with accounting and bookkeeping while it is just more on recording data but plays a crucial role in accounting in terms of gathering the data for financial report, interpretation and analysis. That also concludes that in this result, it garnered similarity percentage of 15.60% which in sense not all applicants with bookkeeping skills can apply to a higher position with certain skills and licensure exams needed like accountancy. One pair also shows above 50% similarity score which is ‘manager’ and ‘leadership’ that makes sense with similar capabilities and roles. This observation also makes sense in ‘Brand Management’ and ‘Product Marketing’ (70.55%) and ‘Sales’ and ‘Presentation’ (78.77%). Some words cannot be found on the datasets and will be considered for improvement recommendations.