**Test Cases:**

Inverted Index Creation and TF-IDF scores calculation for Testing sample:

Document 1: The game of life is a game of everlasting learning

Document 2: The unexamined life is not worth living

Document 3: Never stop learning

Actual Calculated Indexing after processing the documents:

Manual Result {'game': {'1': [1, 3]}, 'life': {'1': [2], '2': [2]}, 'everlast': {'1': [4]}, 'learn': {'1': [5], '3': [3]}, 'unexamin': {'2': [1]}, 'worth': {'2': [3]}, 'live': {'2': [4]}, 'never': {'3': [1]}, 'stop': {'3': [2]}}

Manual TF-IDF Result {'1': {'game': 0.590848501887865, 'life': **0.2352182518111363**, 'everlast': 0.2954242509439325, 'learn': **0.2352182518111363**}, '2': {'unexamin': 0.3692803136799156, 'life': **0.29402281476392034**, 'worth': 0.3692803136799156, 'live': 0.3692803136799156}, '3': {'never': 0.4923737515732208, 'stop': 0.4923737515732208, 'learn': **0.3920304196852271**}}

Query Processing and Boolean and Vector Model results for **“life and learning”**

**Life = Normalized TF(life in query) \* IDF(life in Manual Result) = 0.5\*1.1760912591 = 0.5880456295**

**Learn = NormalizedTF(learn in query) \* IDF(learn in Manual Result) = 0.5\*1.1760912591 = 0.5880456295**

**Query doc1 doc2 doc3**

**Life 0.5880456295 0.2352182518111363 0.29402281476392034 0**

**Learn 0.5880456295 0.2352182518111363 0 0.3920304196852271**

Cosine Similarity (query, d1) = Dot product(query, d1) / ||query|| \* ||d1||

Dot product (query, d1) = query[0] \* d1[0] + query[1] \* d1[1] \* … \* query[n] \* d1[n]

* (**0.5880456295 \* 0.2352182518111363** ) + (**0.5880456295 \* 0.2352182518111363**) =

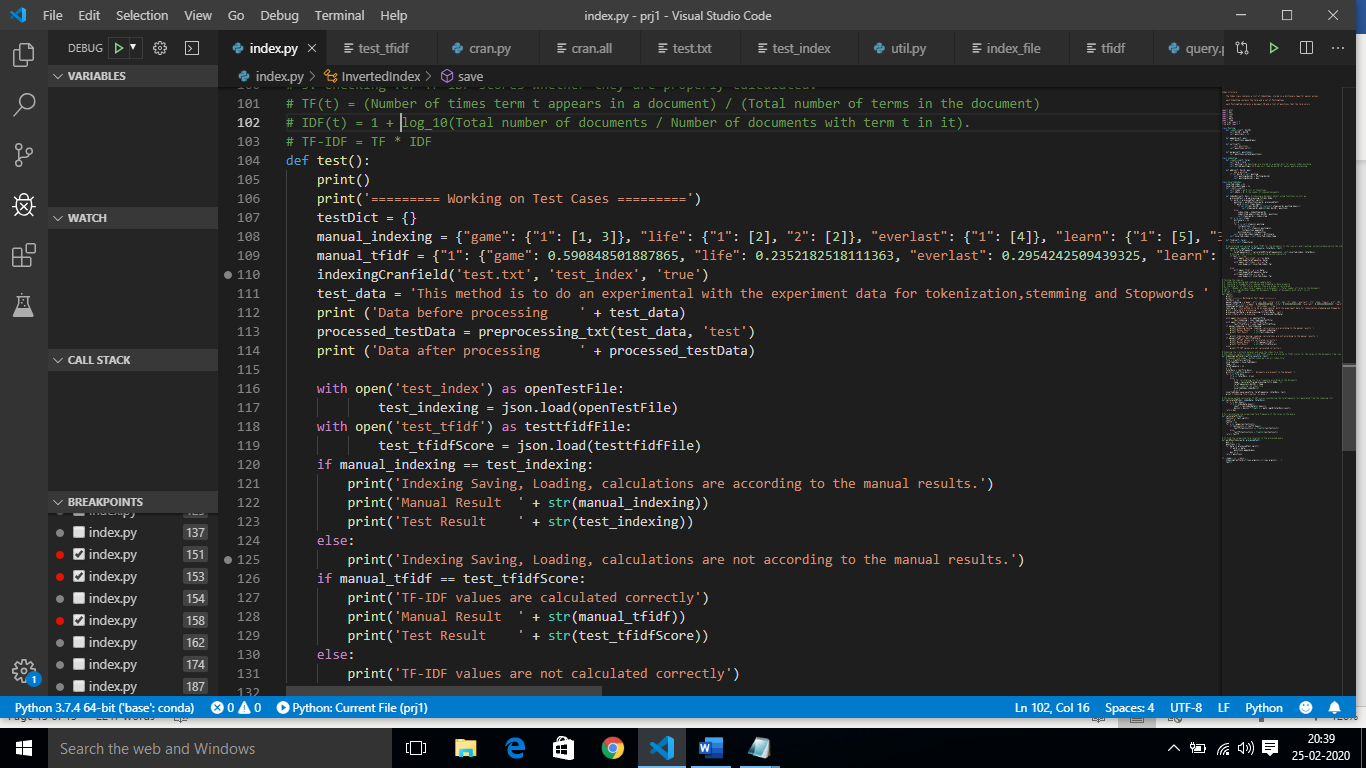
||query|| = square root(query[0]2 + query[1]2 + ... + query[n]2) = square root(**0.5880456295\*\*2+0.5880456295\*\*2) =**  square root(0.6915953248) = **0.8316221046**

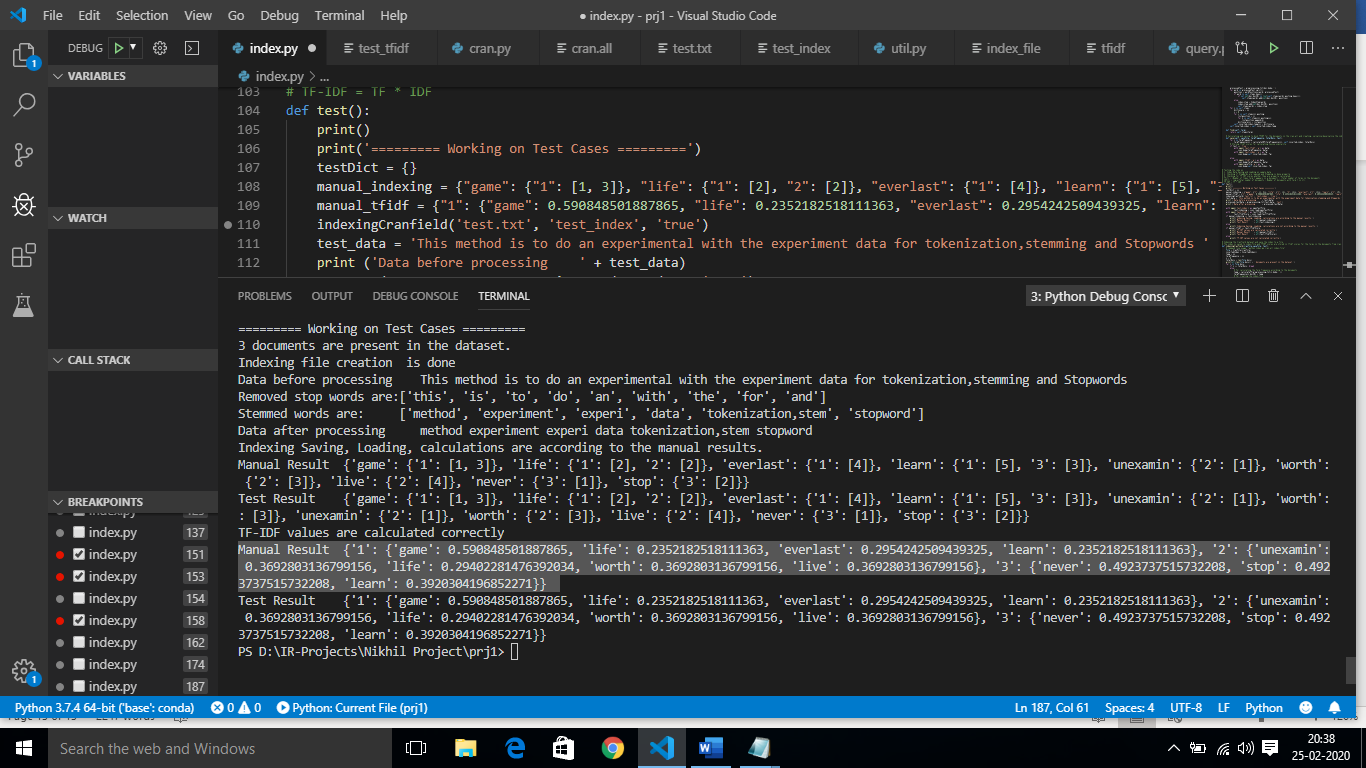
||d1|| = square root(d1[0]2 + d1[1]2 + ... + d1[n]2) = square root(**0.2352182518111363\*\*2 \* 0.2352182518111363\*\*2** ) = square root(**0.11065525196) = 0.33264884181**

Cosine Similarity (query, d1) = **(0.2766381299)/(0.27663812991) = 1**

Cosine Similarity (query, d1) **=** 0.7071067811865476

Cosine Similarity (query, d1) = 0.7071067811865475

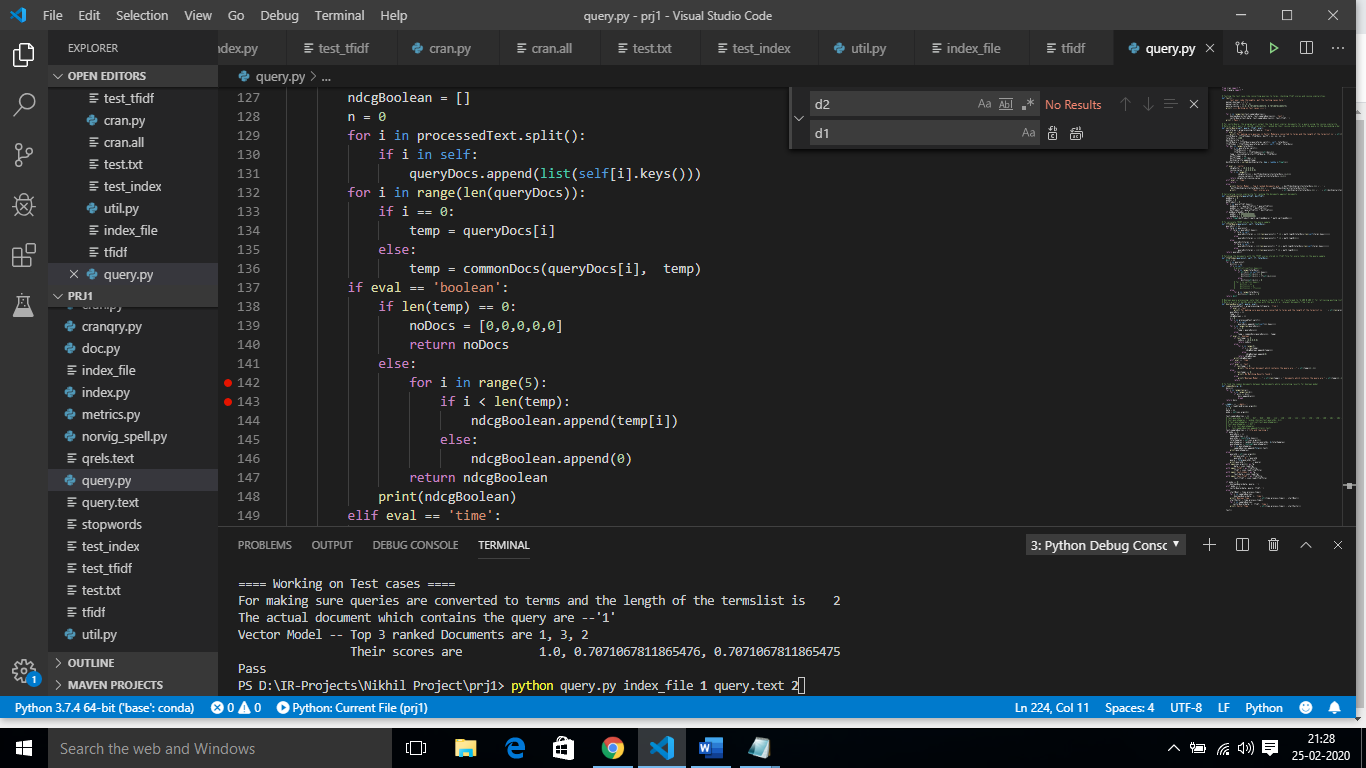




Boolean Result: 1

Vector Model Result: 1, 3, 2

Cosine Similarity: 1.0, 0.7071067811865476, 0.7071067811865475



Actual Data to be fetched for query life and learning is : 1,3

Boolean Model = NDGC\_5(y\_true, y\_score) = NDGC([1,0,0,0,0] + [1,0,0,0,0]) = 1.0

|  |  |  |
| --- | --- | --- |
| docId | y-true | Y\_score |
| 1 | 1 | 1 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |

Vector Model = NDGC\_5(y\_true, y\_score) = NDGC([1,0,0,0,0] + [1.0,0.7071067811865476, 0.7071067811865475,0,0]) = 0.6934264036172708

|  |  |  |
| --- | --- | --- |
| docId | y-true | Y\_score |
| 1 | 1 | 1.0 |
| 3 | 1 | 0.7071067811865476 |
| 2 | 0 | 0.7071067811865475 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |

