

National Skill Development Agency (NSDA)

MYAS NIC Cell

Kaushal Bhawan, New Delhi

Date: 19/01/2018

Project: A Decision Support System for NSQF Proposals for Acceptance and Introduction of New Qualification Packs and Maintenance of Miscellaneous Information and Work Flow Components.

Meeting Agenda: Discussions on Solution Architecture for the Proposed System, Chaired by DG (NSDA)

Agenda Point: Machine Learning Techniques based NLP for Resolving Duplicate / Fraudulent / Mala Fide / Deceptive Proposals

I. Corpus Sample Proposal: Variance Sampling of the Proposals Submitted in “**Similar**” Category

- (i) Doc 1: NSQF-Light Motor Vehicle Driver.pdf
- (ii) Doc 2: NSQF-Taxi Driver.pdf
- (iii) Doc 3: QF - LMV Owner Driver.pdf

Table-I Similarity / Variance Matrix

Experiment No.	Sample Source Document ID	Vs.	Sample Destination Document ID	Similarity / Variance Score %
1	Doc1	Vs.	Doc 2	0.9501619431737199
2	Doc1	Vs.	Doc3	0.3719246949784795
3	Doc2	Vs.	Doc3	0.4140482891413597

II. Corpus Sample Proposal: Variance Sampling of the Proposals Submitted in “**Different**” Categories

- (i) Doc 1: QF _Accounting_BAN101.pdf
- (ii) Doc 2: NSQF- Advance Pattern Maker (CAD-CAM).pdf
- (iii) Doc 3: NSQF-2-3 WHEELERS MECHANIC TECHNICIAN (revised on 24.1.2017).pdf

Table-II Similarity / Variance Matrix

Experiment No.	Sample Source Document ID	Vs.	Sample Destination Document ID	Similarity / Variance Score %
1	Doc1	Vs.	Doc 2	0.6710649426110178
2	Doc1	Vs.	Doc3	0.448465462510357
3	Doc2	Vs.	Doc3	0.3848116334975575