REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Deployment of AI-Enabled Automated Solution at AWS Cloud

Name of the Presenter

SRN: Gaurav Karki

Date: 27th Aug 2022

M.Tech in Artificial Intelligence

Capstone Project Presentation
Year: II

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Background | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Technical/Functional Problem

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Resources Specification
Software | Hardware

07 Software Design
High Level | Low Level Design Diagrams

08 Implementation
Demo | Application | Use cases

09 Testing and Validation
Test Results | Learnings

10 Analysis and Results
Key Findings | Insights

12 Suggestions and Conclusion
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

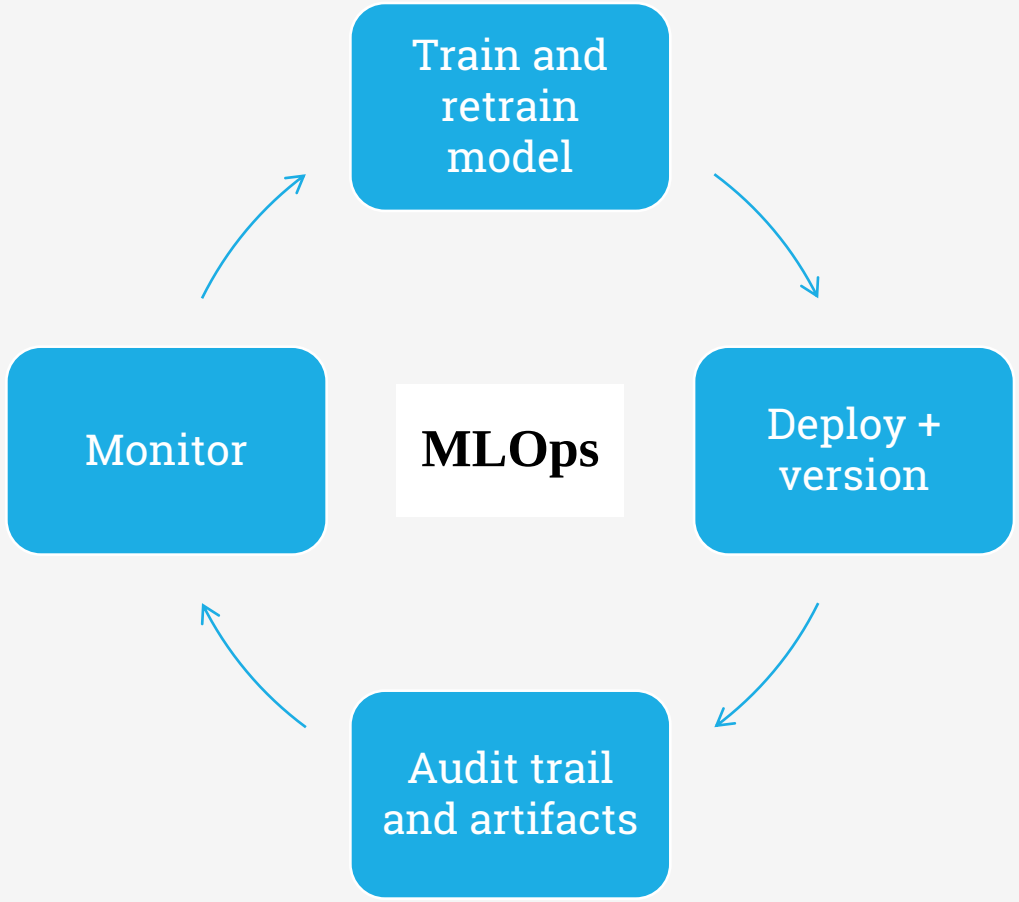
Background | Current status | Why this topic

Machine Learning Ops(MLOps)

MLOps helps enterprises improve AI ROI by automating and scaling the machine learning lifecycle.

AWS Cloud Service

Using cloud services allows you to create and deliver ML models without developing MLOPs solutions.



Introduction

Background | Current status | Why this topic

ML Hierarchy

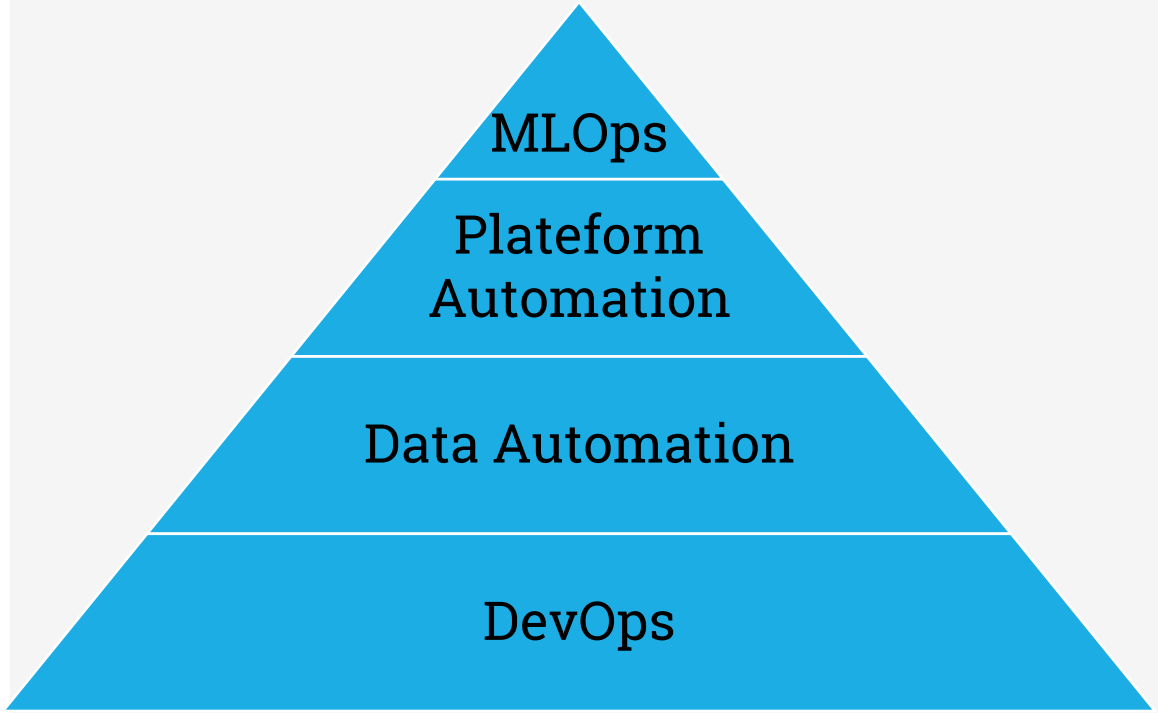


Figure: ML Engineering Hierarchy of Needs
Source: https://www.smartsheet.com/content/utilization-management

Challenges with MLOps

- Complexities with Data
- Engineering and Deployment
- Integration Risks

Reason

- Manual Process

Opportunity

- AWS cloud service for MLOps

Literature Review

Seminal works | Summary | Research Gap

Title	Author & Year	Journal/Source	Major Insights	Research Gap
Measuring progress in improving prior authorization	American Medical Association | 2021	https://www.ama-assn.org/system/files/prior-authorization-reform-progress-update.pdf	• 84% of physicians say the prior authorization burden is “high” or “extremely high”.	Streamlining PA Process
CAQH Index Report	The Council for Affordable Quality Healthcare | 2020	https://www.caqh.org/sites/default/files/explorations/index/2020-caqh-index.pdf	• Physicians do 40 PAs every week on average. • PA takes 20 minutes to accomplish manually. • Automating PA might save $417 million yearly.	Use of Automation

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Literature Review

Seminal works | Summary | Research Gap

Title	Author & Year	Journal/Source	Major Insights	Research Gap
Why Should You Use MLOps? - Amazon SageMaker.	AWS | 2022	https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-projects-why.html	• AWS provide inbuilt services and cloud platform	Development is required
Rule-based prediction of medical claims	J. Wojtusiak, C. Ngufor, J. Shiver, and R. Ewald | 2011	Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011	• Attributional techniques can predict claim discrepancies.	Fraud detection differs from claims rework review.
A text similarity approach for precedence retrieval from legal documents	Thenmozhi, D Kannan, C Aravindan | 2022	ceur-ws.org	• Word embedding such as word2vec captures document semantics.	Semantics information at sentence level

Problem Statement

Technical | Functional

Technical

- To find text similarity in the patient's clinical history with the insurer's guidelines to approve a PA claim.
- To deploy ML model using ASW Lambda

Functional

- To streamline the PA process to prevent delays in patient treatment and decrease costs for insurers and providers.
- Many ML projects fail to meet their objectives because automating and operationalizing ML solutions is so challenging

Project Objectives

Primary & Secondary Objectives | Expected Outcome

Objective

- To develop an AI-enabled solution, consisting of a text similarity model based on NLP and decisions based on semantic analysis, to make utilization management a streamlined and collaborative process.
- To deploy solution from the previous study using the cloud service AWS and examines the possibility for automating and operationalizing manual ML operations

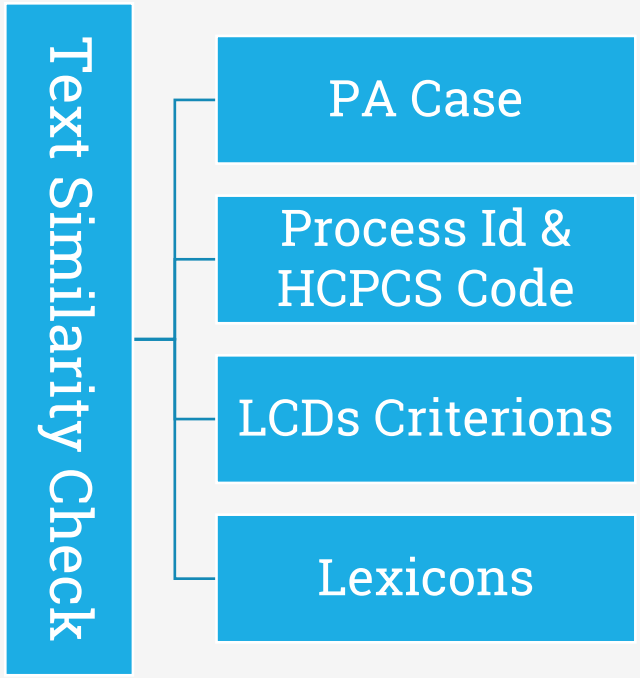
Sub-Objective

- Automate PA to improve auto-decisions and prioritize clinical evaluations.
- To expedite patient care while increasing member satisfaction.
- Unlocking unstructured data to contextualize permission requests can improve insurers' administrative efficiency and provider experience.
- Monitoring and logging

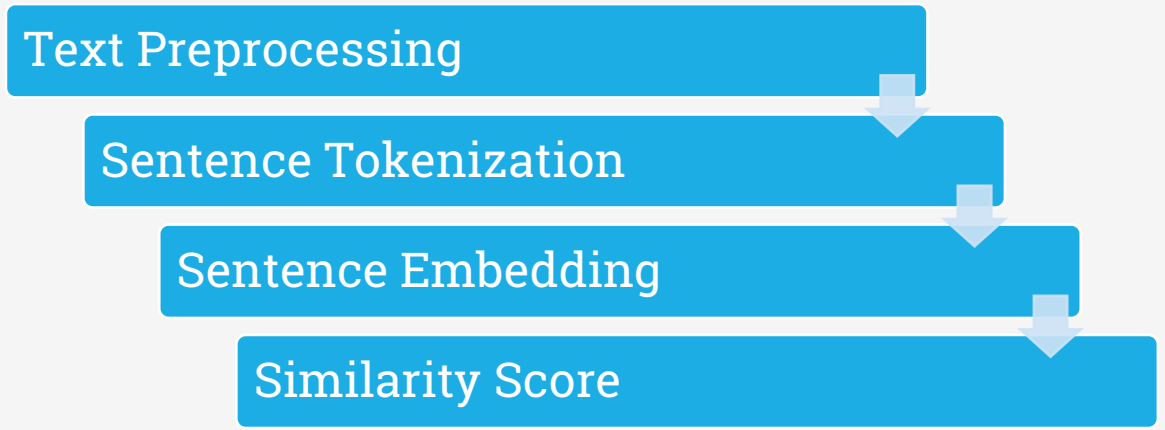
Project Methodology

Conceptual Framework | Research Design

Conceptual Framework



Text Analysis Workflow



Project Methodology

Conceptual Framework

Figure No. Maturity Model

Resource Specifications

Software | Hardware | Others

Data Resources

- **Process**
  Powered Mobility Devices
- **Guidelines**
  Local Coverage Determination, which are decisions made by a Medicare Administrative Contractor (MAC).
- **PA Case**
  Powered Mobility Devices

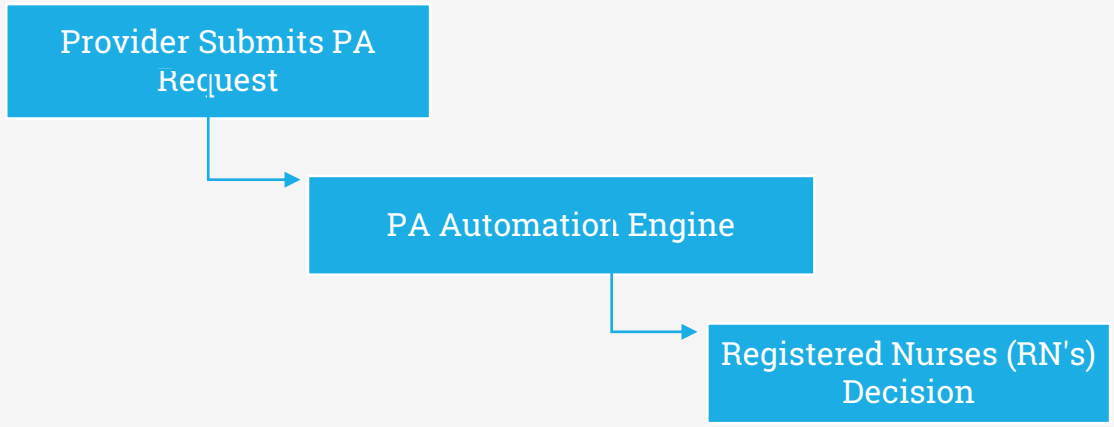
Technical Resources

- Python environment with required libraries like NLTK, Pandas, Tensorflow, Scipy etc.
- Semantic Textual Similarity (STS) benchmark data for evaluation.
- Pre-trained model - Universal Sentence Encoder
- AWS CLI
- AWS management console

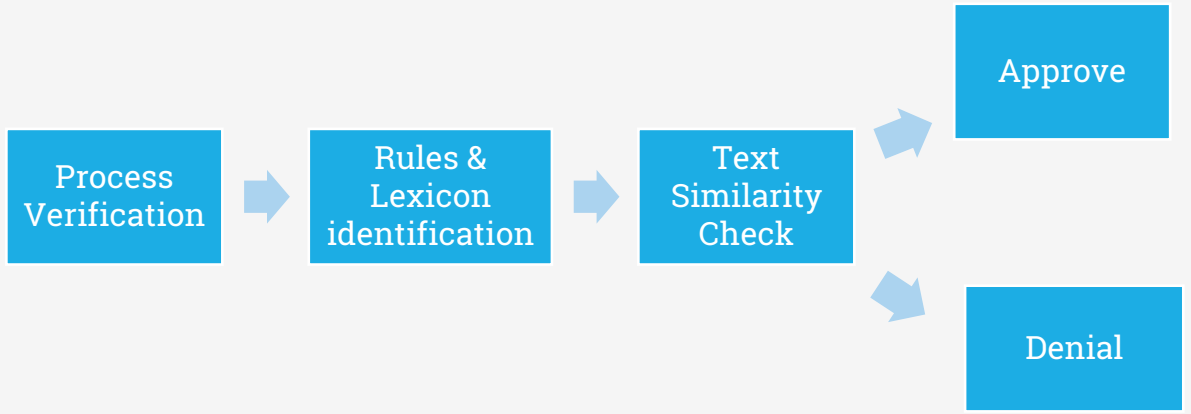
Software Design

High | Low Level Designs

Proposed AI-Enabled PA Process



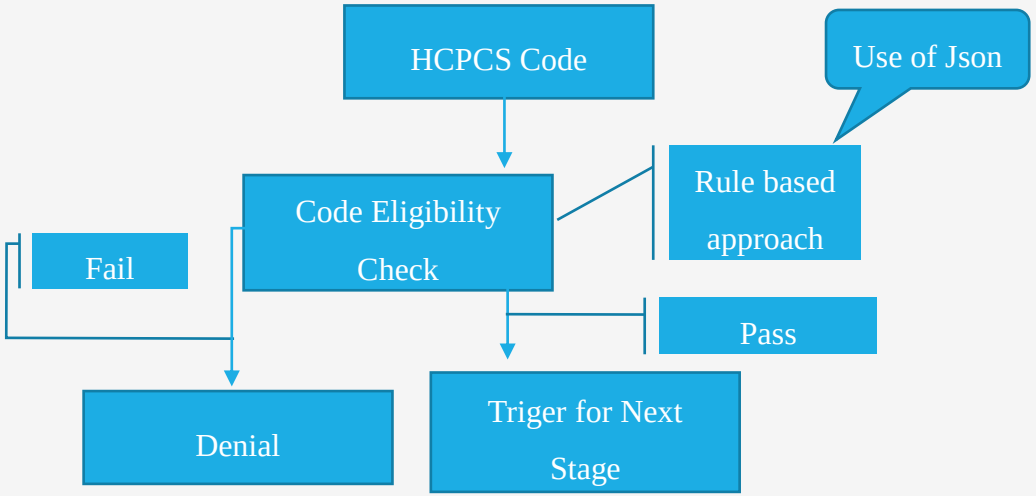
Automation Engine Process



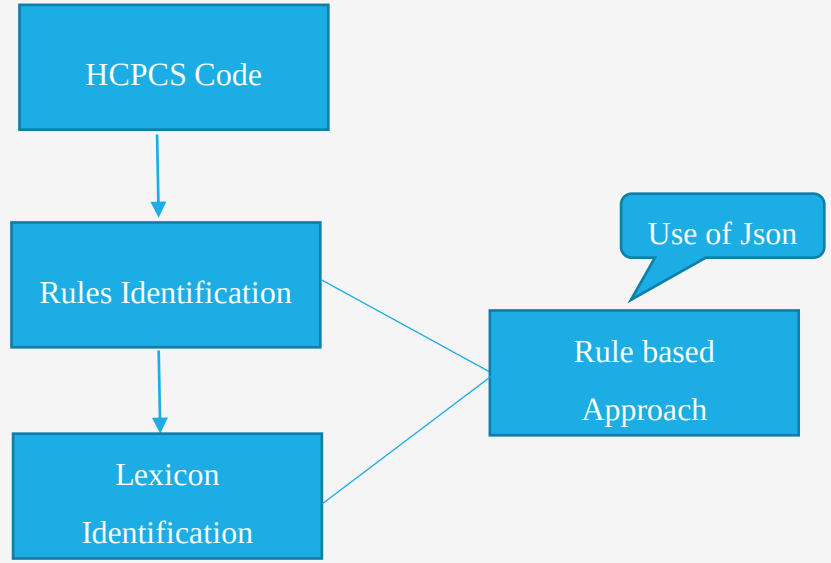
Software Design

High | Low Level Designs

Process Verification



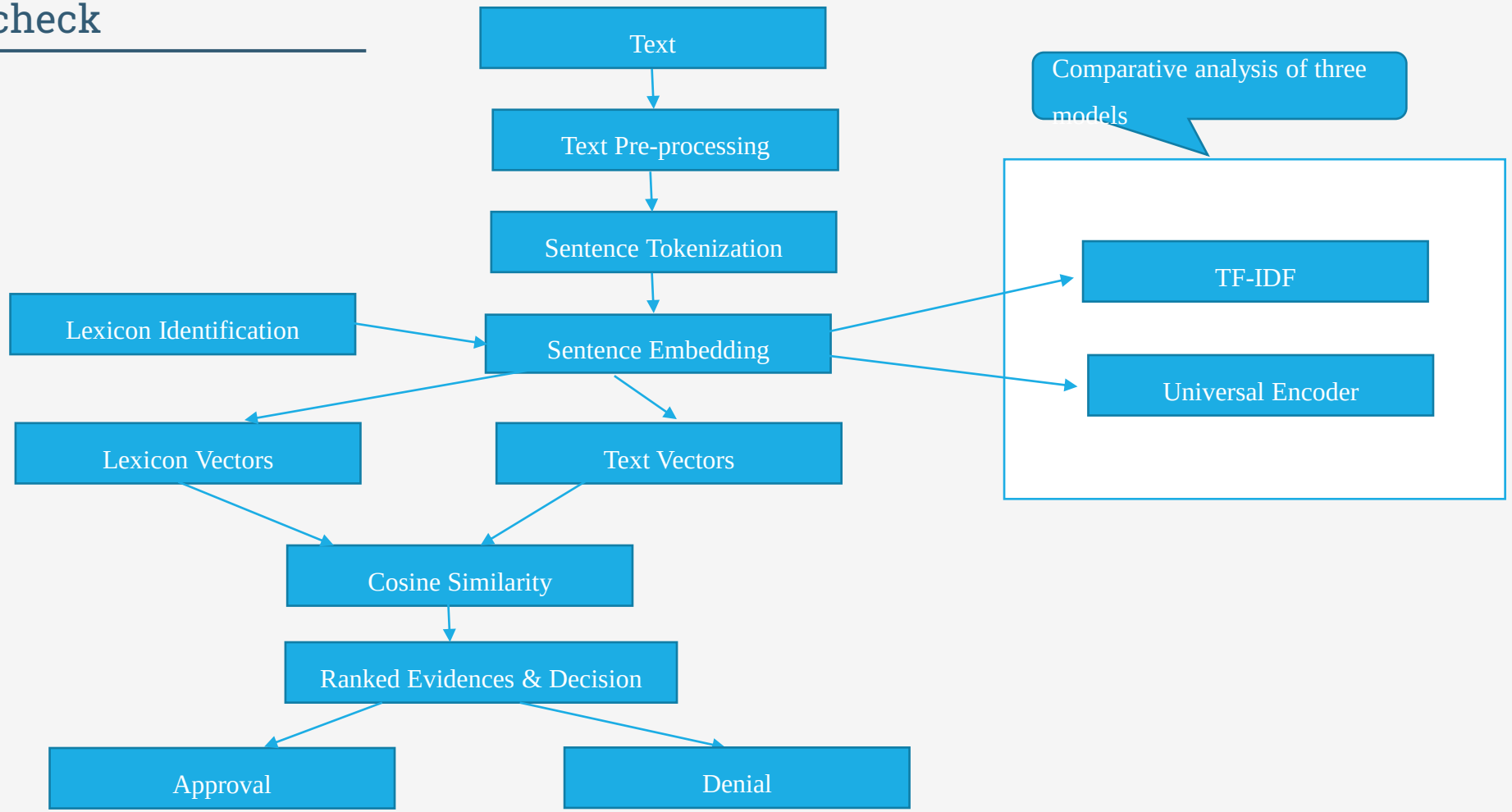
Rules and Lexicon Identification



Software Design

High | Low Level Designs

Text Similarity check



AWS MLOps Design Overview

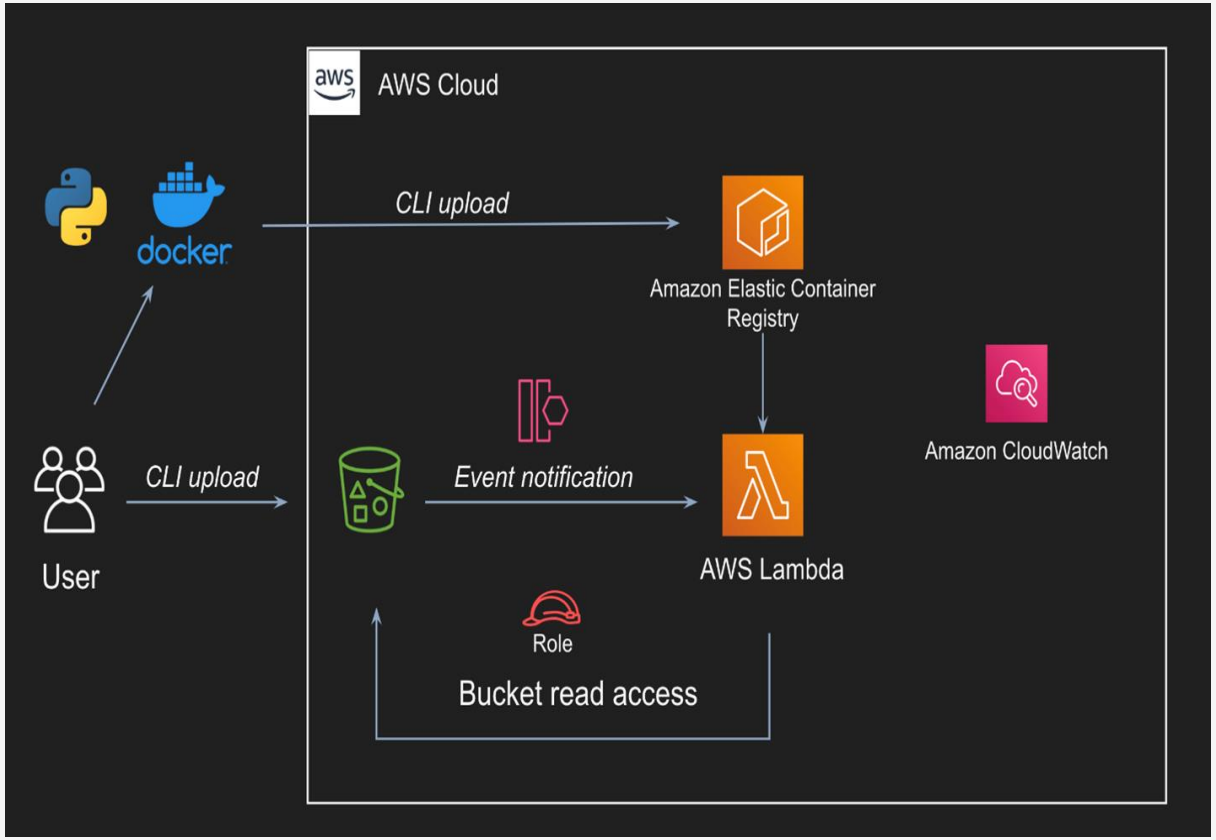


Figure No. Deployment Process with Each AWS Service Connection

Implementation

Demo | Application | Use cases

DEMO

Testing and Validation

Test Results | Learnings

Testing for TF_IDF Technique

Rule name	Total no. of Matches	Top matching sentence from PA text	Highest score
A	5	he limited in his ability to participate in all mobility related activities of daily living in the home setting	0.491431
B	6	he la unable to safely or effectively use cane or walker for the distance needed in the home due to fatigue joint pain and numbness	0.2933
C	9	he is unable to self-propel an optimally configures manual wheelchair due to upper extremity weakness and arthritic hand pain	0.3841

Testing for Universal Sentence Encoder Technique

Rule name	Total no. of Matches	Top matching sentence from PA text	Highest score
A	7	he limited in his ability to participate in all mobility related activities of daily living in the home setting	0.72
B	5	he la unable to safely or effectively use cane or walker for the distance needed in the home due to fatigue joint pain and numbness in rle	0.70
C	8	he is unable to self-propel an optimally configures manual wheelchair due to upper extremity weakness and arthritic hand pain	0.71

Testing and Validation

Test Results | Learnings

Testing for Universal Sentence Encoder

2022-08-26T20:24:33.032+05:30	loading Universal sentence encoder
2022-08-26T20:24:34.965+05:30	2022-08-26 14:54:34.965469: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could…
2022-08-26T20:24:34.965+05:30	2022-08-26 14:54:34.965509: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to …
2022-08-26T20:24:34.965+05:30	2022-08-26 14:54:34.965537: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel dri…
2022-08-26T20:24:34.965+05:30	2022-08-26 14:54:34.965914: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow …
2022-08-26T20:24:34.965+05:30	To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2022-08-26T20:24:39.261+05:30	START RequestId: 50dcd3fd-d3d2-4c98-8218-460ae8962bac Version: $LATEST
2022-08-26T20:24:39.331+05:30	OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k
2022-08-26T20:24:39.921+05:30	2022-08-26 14:54:39.921591: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could…
2022-08-26T20:24:39.921+05:30	2022-08-26 14:54:39.921628: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cud…
2022-08-26T20:24:41.376+05:30	loading Universal sentence encoder
2022-08-26T20:24:43.052+05:30	2022-08-26 14:54:43.052533: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could…
2022-08-26T20:24:43.052+05:30	2022-08-26 14:54:43.052576: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to …
2022-08-26T20:24:43.052+05:30	2022-08-26 14:54:43.052601: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel dri…
2022-08-26T20:24:43.052+05:30	2022-08-26 14:54:43.052852: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow …
2022-08-26T20:24:43.052+05:30	To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2022-08-26T20:24:56.038+05:30	done
2022-08-26T20:24:56.257+05:30	processing file:
2022-08-26T20:24:56.257+05:30	performing ocr....
2022-08-26T20:24:56.276+05:30	applying process rules...
2022-08-26T20:25:04.365+05:30	Put Complete
2022-08-26T20:25:04.369+05:30	END RequestId: 50dcd3fd-d3d2-4c98-8218-460ae8962bac

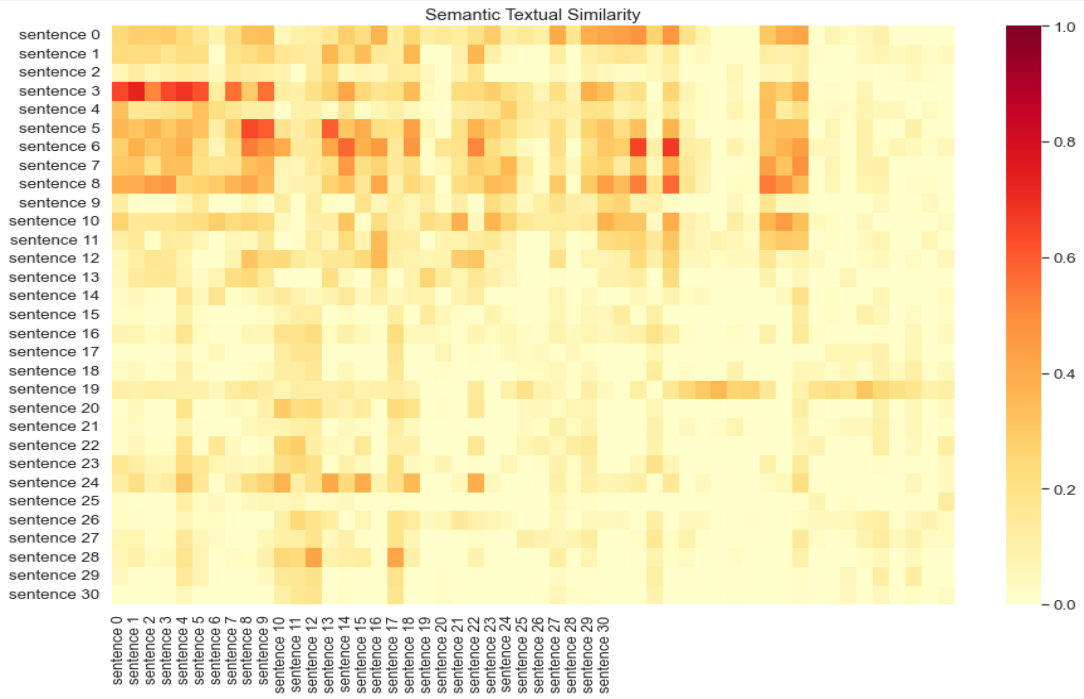
Figure CloudWatch Logs

Analysis and Results

Key Findings | Insights

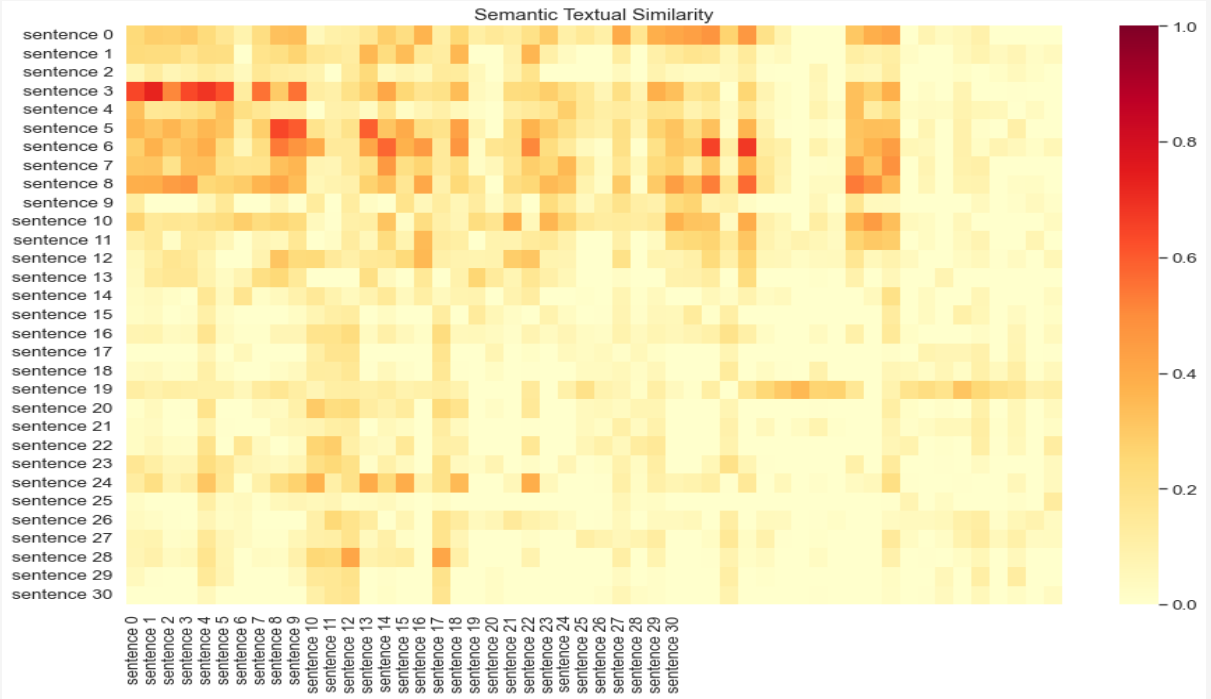
Results for TF_IDF Technique

Pearson correlation coefficient = 0.2340
p-value = $1.015e^{-19}$



Testing for Universal Sentence Encoder Technique

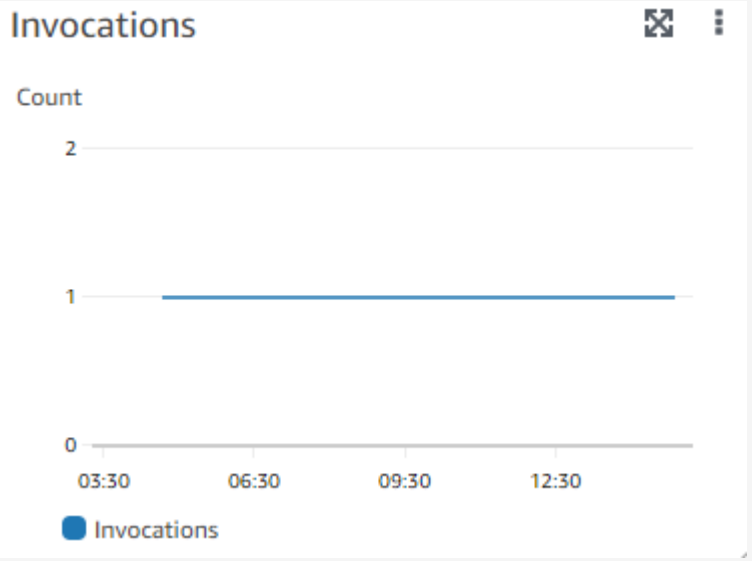
Pearson correlation coefficient = 0.83
p-value = 0

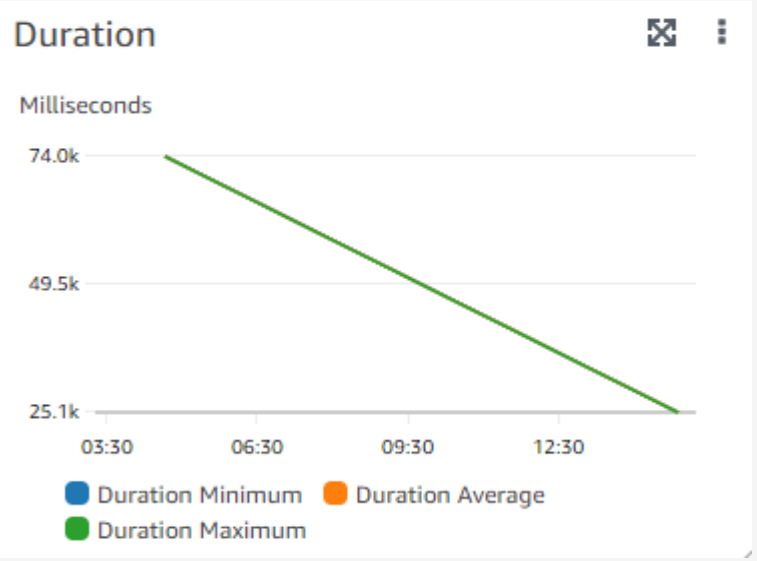


Analysis and Results

Key Findings | Insights

Results from CloudWatch Logs





Suggestions and Conclusion

Insights | Next Step | Future Scope

Insights

- The previous section shows that USE's techniques are superior to TF IDF's.
- Sentence embedding approaches represent whole sentences as vectors. This helps the machine recognize context, intention, and other complexities.
- Fewer denied claims, lower costs, more effective treatments, real information, and more efficient resource use benefit the healthcare business.

Next Step

- As this study only examines a powered wheel chair process, it is necessary to validate this method with a more intricate procedure.
- The second step for insurers is to ask employees about their prior authorization experiences.

Future Scope

- This study motivates us to solve this problem by approaching it as a classification problem.
- Termination Policy – API response consistent will not go for loop
- Python Profiling – Performance can get increase from less resources

References

Bibliography | Webliography

[1] T. M. Wickizer and D. Lessler, "Utilization management: Issues, effects, and future prospects," *Annu. Rev. Public Health*, vol. 23, pp. 233–254, 2002, doi: 10.1146/ANNUREV.PUBLHEALTH.23.100901.140529.

[3] A. Medical Association, "Prior Authorization Physician Survey Update | AMA," 2022, Accessed: Aug. 10, 2022. [Online]. Available: https://www.ama-assn.org/system/files/prior-.

[4] "Most physicians had little relief from prior authorization as COVID cases soared | American Medical Association." https://www.ama-assn.org/press-center/press-releases/most-physicians-had-little-relief-prior-authorization-covid-cases (accessed Aug. 10, 2022).

[5] "Electronic Prior Authorizations Impact Quality, Provider Burden." https://healthpayerintelligence.com/news/electronic-prior-authorizations-impact-quality-provider-burden (accessed Aug. 10, 2022).

[6] M. Kumar, R. Ghani, and Z. S. Mei, "Data mining to predict and prevent errors in health insurance claims processing," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 65–73, 2010, doi: 10.1145/1835804.1835816.

[7] J. Wojtusiak, C. Ngufor, J. Shiver, and R. Ewald, "Rule-based prediction of medical claims' payments: A method and initial application to medicaid data," *Proc. - 10th Int. Conf. Mach. Learn. Appl. ICMLA 2011*, vol. 2, pp. 162–167, 2011, doi: 10.1109/ICMLA.2011.126.

[8] D. Thenmozhi, K. Kannan, C. A.-F. (Working Notes), and undefined 2017, "A text similarity approach for precedence retrieval from legal documents.," *ceur-ws.org*, Accessed: Aug. 10, 2022. [Online]. Available: http://ceur-ws.org/Vol-2036/T3-9.pdf.

Annexure

Publications | Conferences

Deployment of AI-Enabled Automated Solution at AWS Cloud

ORIGINALITY REPORT

7%	6%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.oreilly.com Internet Source	4%
2	arxiv.org Internet Source	1%
3	Submitted to National College of Ireland Student Paper	1%
4	docs.aws.amazon.com Internet Source	1%
5	Vinay Khedekar, Yun Tian. "Multi-Tenant Big Data Analytics on AWS Cloud Platform", 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020 Publication	<1%
6	hub.apitree.com Internet Source	<1%
7	Seyedehnafiseh Mirniaharikandehei, Joshua VanOsdol, Morteza Heidari, Gopichandh Danala et al. "Developing a Quantitative Ultrasound Image Feature Analysis Scheme to	<1%

	Internet Source	<1%
8	www.cio.com Internet Source	<1%
9	www.ama-assn.org Internet Source	<1%
10	scholarbank.nus.edu.sg Internet Source	<1%
11	Submitted to Lovely Professional University Student Paper	<1%
12	3015elementcenter.com Internet Source	<1%
13	Submitted to Nanyang Technological University Student Paper	<1%
14	Submitted to University College London Student Paper	<1%
15	data.dws.informatik.uni-mannheim.de Internet Source	<1%
16	www.cookchp.org Internet Source	<1%

Exclude quotes On Exclude matches < 10 words

REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Thank You!