STATATHON 2025

TITLE PAGE

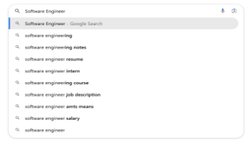
- **Problem Statement ID:** 5
- **Problem Statement Title:** AI-enabled semantic search for National Classification of Occupation (NCO)
- **PS Category- Software/Hardware:** Software
- **Team ID:** 3499
- **Team Name (Registered on portal):** Data Envisors

Data Envisors

IDEA TITLE

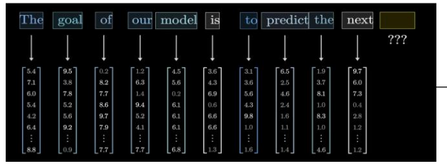
STATATHON 2025

PROBLEM STATEMENT - DESCRIPTION

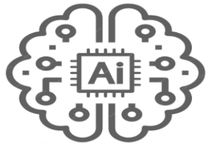


User enters a natural-language search query describing the desired job or skills.

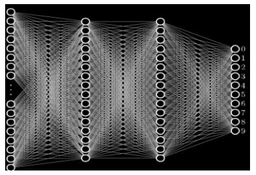
- Current NCO search is keyword-based and requires users to know the exact occupation title.
- This limits discoverability of related jobs, as synonyms and context are ignored.



Converted to Vector representation



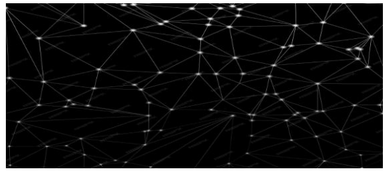
Processing the Query



Outputting Numbers

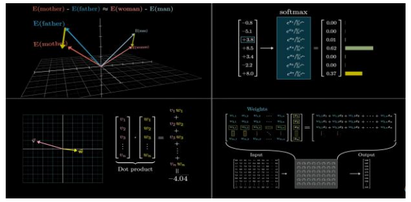
HIVE : DESCRIPTION

- An AI-powered semantic search engine that understands meaning, not just words.
- Uses natural language processing and vector embeddings to match user queries with relevant occupations in the NCO dataset.
- Improves access to job information for seekers, recruiters, and policymakers.

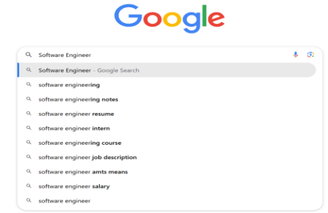


Vector Matching Engine

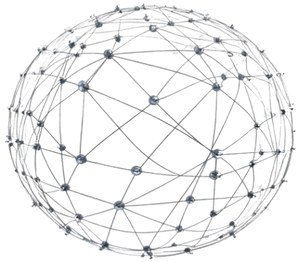
Sematic Search



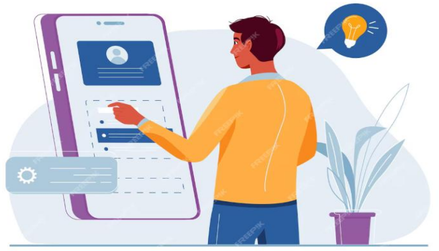
Top N Matching Occupations



INNOVATION AND UNIQUENESS



- **Context-aware search:** Retrieves jobs even when different terminology is used.
- **Smart ranking:** Prioritizes the most relevant occupations based on semantic similarity.
- **Extensible design:** Can be expanded to multilingual search and other datasets.



Data Envisors

TECHNICAL APPROACH

STATATHON 2025

Data Source & Storage → **Embedding & Processing** → **API & Search** → **Frontend Display**

WORK FLOW

React.js UI

React JS

CLOUD

aws Azure

Search Results

Google

Output Stage

PROJECT TECH STACK

pandas

SBERT

Sentence **Transformers**

Embedding Model

NLP

NLP Pipeline

React JS

tailwindcss

Frontend

NCO

DATASET

Pinecone

Vector Database

DATA Stack

Programming language

FastAPI

Framework for APIs

Data Processing & Semantic Search stage

docker + Azure

Cloud hosting
&
Containerization tool

Data Envisors

FEASIBILITY AND VIABILITY

STATATHON 2025

FEASIBILITY

1. Technical Feasibility

- Pre-trained embedding models (all-MiniLM-L6-v2 / text-embedding-3-large) ready for use
- FAISS / Pinecone integration with FastAPI for efficient vector search

FastAPI

Pinecone

2. Operational Feasibility

- User-friendly React.js interface for non-technical users
- Fast, accurate results through optimized semantic search

docker

3. Resource Feasibility

- Open-source tools minimize licensing costs
- NCO dataset eliminates need for custom data collection

Azure

4. Time Feasibility

- Pre-built libraries accelerate development
- Modular design enables parallel frontend & backend work

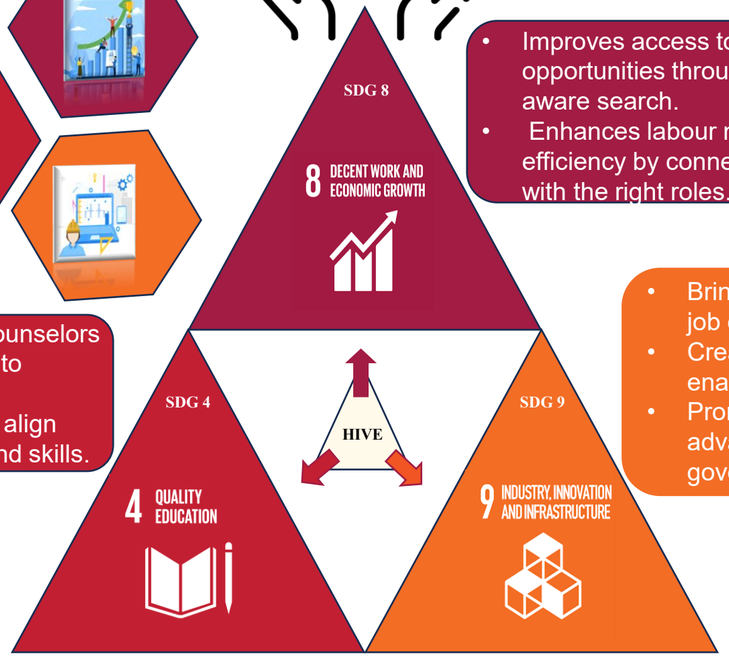
VIABILITY

Alignment with UN SDGs ensures long-term social, economic, and environmental relevance.

UNDP

SUSTAINABLE DEVELOPMENT GOALS

- Improves access to job opportunities through context-aware search.
- Enhances labour market efficiency by connecting talent with the right roles.



- Helps students and counselors link skills and training to occupations.
- Guides institutions to align courses with in-demand skills.

- Brings AI innovation to public job databases.
- Creates a scalable, tech-enabled search system.
- Promotes adoption of advanced search technology in government platforms.

Data Envisors

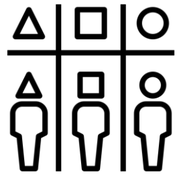
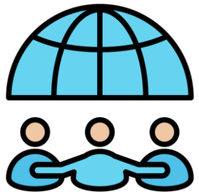
IMPACT AND BENEFITS

STATATHON 2025

BENEFITS

- Improved Job Accessibility
- Increased Employment Opportunities
- Convenience and Speed
- Skill-to-Job Mapping
- Inclusive Access
- Better Career Guidance
- Government Data Utilization
- Scalable and Future-Ready

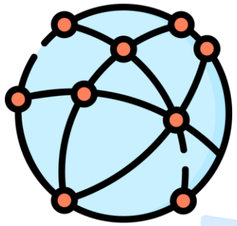
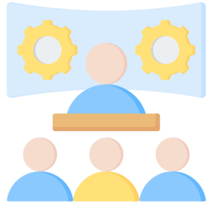



MARKET DEMAND & IMPACT

- Government Adoption
- Job Portals Integration
- Career Counseling Platforms
- Expansion to Other Domains
- Skill Development Programs
- AI-powered Analytics



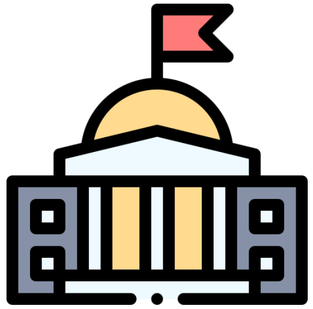




Data Envisors

RESEARCH AND REFERENCES

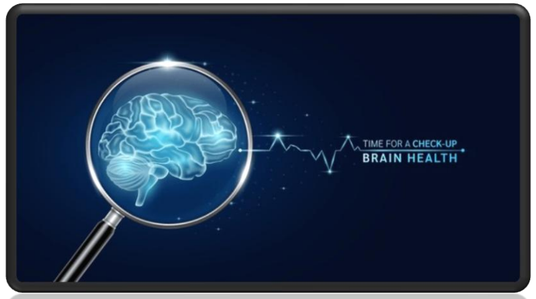
STATATHON 2025

1. National Classification of Occupations (NCO) 2015 – Ministry of Labour & Employment
Official dataset containing job codes, titles, and descriptions.
https://labour.gov.in



2. Semantic Search in AI

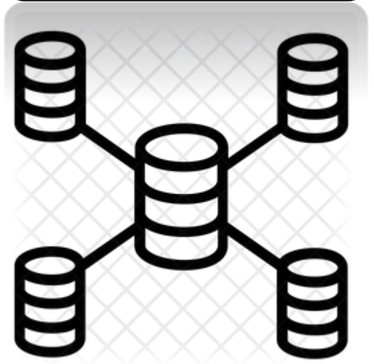
- "Semantic Search: The Future of Information Retrieval" – IEEE Xplore
- Explains how embeddings and vector databases improve contextual search.



3. Vector Databases
FAISS (Facebook AI Similarity Search) – Open-source library for efficient similarity search.
https://faiss.ai

4. Embedding Models
Sentence Transformers Documentation – For generating text embeddings.
https://www.sbert.net

5. Government Data Utilization
World Bank Report on the use of open data for economic growth.



6. Related Hackathon Problem Statements
Past use cases in data dissemination and job-matching AI tools from similar competitions.