

STATATHON 2025



- Problem Statement ID: 5
- Problem Statement Title: Al-enabled Semantic Search for National Classification of Occupation (NCO)
- PS Category: Software
- Team ID: 5524
- Team Name (Registered on portal): Penguins



DAsort



Solution: An AI-powered, multilingual semantic search engine for NCO-2015 that enables fast, accurate occupation code retrieval using natural language, voice input, and synonym/context understanding.

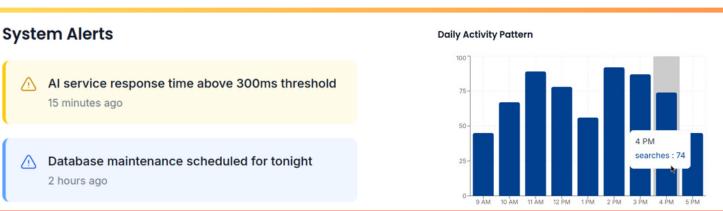
Detailed Explanation

- Convert NCO-2015 dataset into structured JSON/CSV preserving hierarchy.
- Use NLP embeddings (BERT/Sentence Transformers) for semantic understanding.
- Rank search results by relevance & confidence scores.
- Support synonyms, spelling variations, and related term mapping.
- Admin panel for database updates & search audit logs.

#1 96% match Elected Official, State Government Elected Official, State Government serves in various capacities in legislative and executive branches of State Government. Performs, administrative, executive or advisory functions, which include leading council of ministers of State Government; shaping, determining and defining government policies, and directing application and execution of government policies through departmental heads; presiding over deliberations of Houses of State Legislatures; committees and VOLUME II A National Classifica...

Problem & Innovation

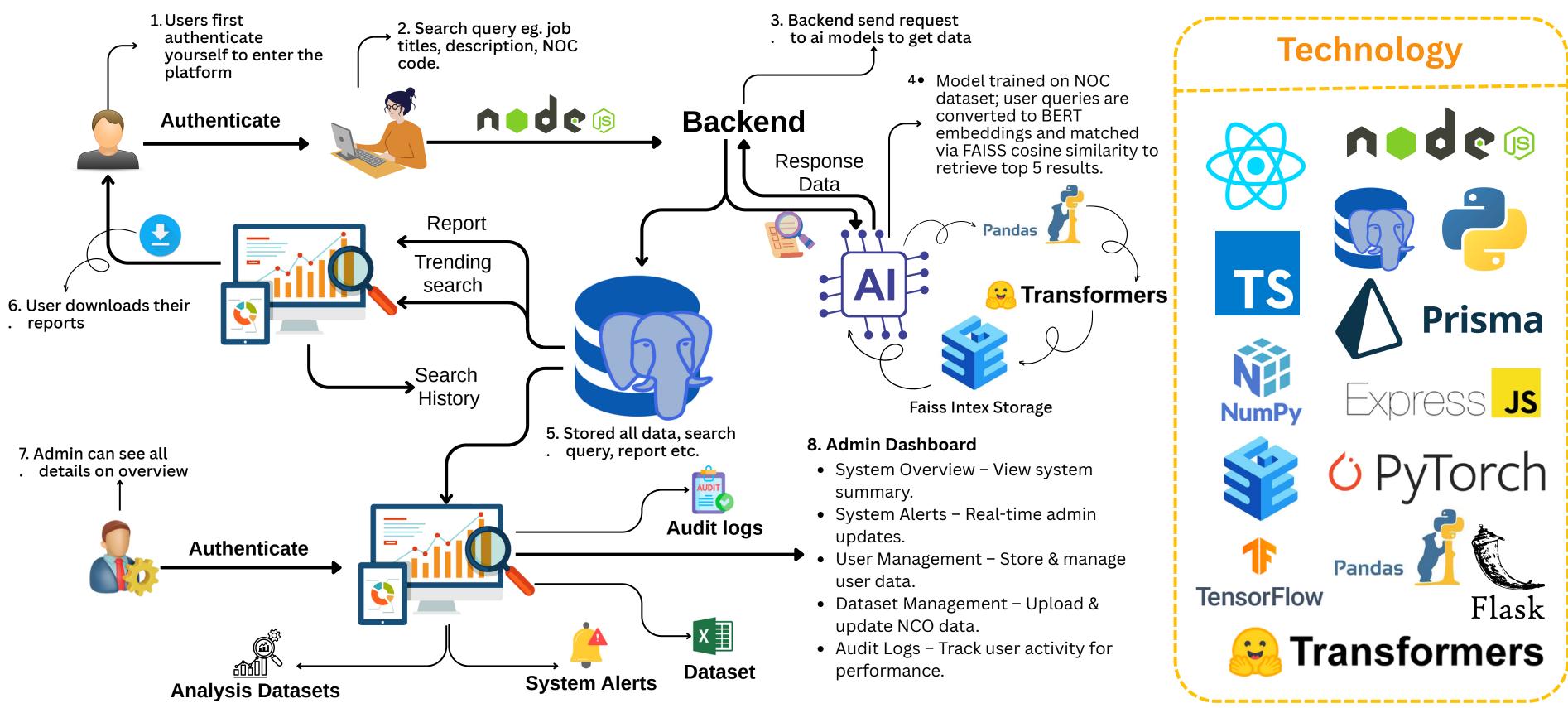
- Eliminates dependency on exact keyword search in PDFs.
- Reduces classification errors & speeds up survey workflows.
- Improves accuracy for 50,000+ enumerators nationwide.
- Multilingual + voice search for real-time field usability.
- Domain-specific synonym/context mapping for Indian occupations.
- API-ready for integration with MoSPI systems.





TECHNICAL APPROACH







FEASIBILITY AND VIABILITY



Feasibility Analysis:



- Uses proven technologies to deliver fast, accurate, and scalable real-time search and classification.
- Easily integrates with existing government systems, supports multilingual and voice input, and ensures security with encryption and role-based access.



- Significantly reduces classification time, works seamlessly across devices, and handles large-scale deployment without disrupting workflows.
- Supports bulk data processing and offers a user-friendly interface that requires minimal training for adoption.



- Cuts down on manual work, delays, and backlogs, saving costs on manpower redistribution and inefficient processes.
- Reduces training expenses and ensures long-term savings through a scalable, lowmaintenance infrastructure.

Challenges & Risks

- Performance & Accuracy:
 Fast searches with high classification accuracy.
- Multilingual & Context Handling:
 Accurate, error-free translations across all languages.
- Security:
 Safeguarding sensitive data and ensuring compliance.

Strategies to Overcome Challenges

- Description of the second o
- Strengthen Multilingual Capabilities
 Get expert-checked translations with context-aware AI.
- Dse secure access, encryption, and audits...



IMPACT AND BENEFITS

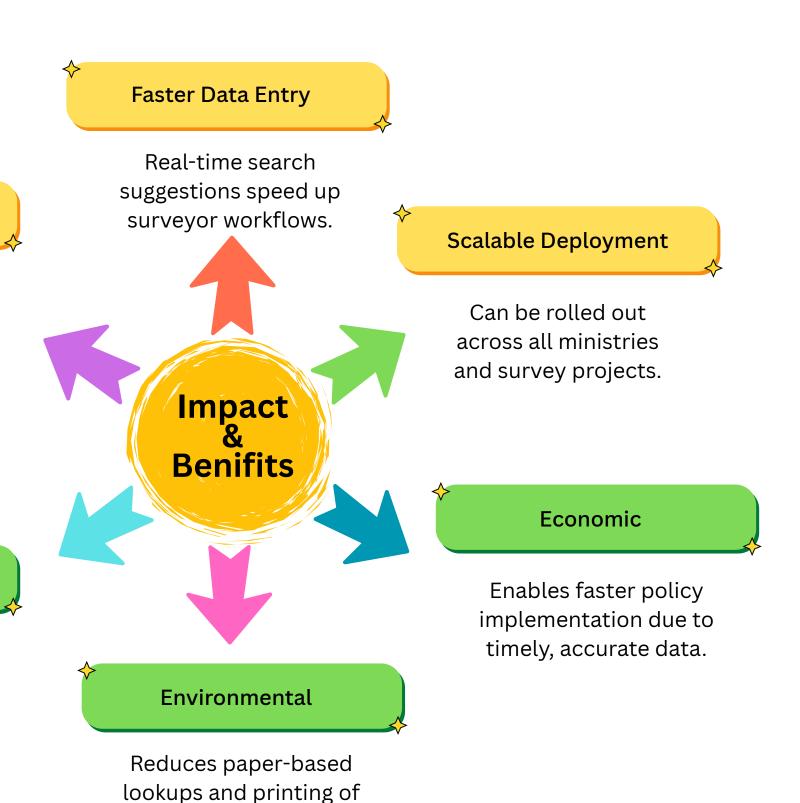


Improved Accuracy in Job Classification

AI-powered semantic matching reduces human error in selecting NCO codes.

Social

Citizens' employment data is recorded more accurately, enabling better policy decisions.



NCO manuals.

Revenue Model

Dual Monetization Approach

- 1. **Government Contracts** (high-ticket, B2G sales)
- 2. SaaS Model for Private Sector (recurring, scalable)
- 3. **Advertising Revenue** (from high-traffic public-facing tools).
- Implementation Contracts: Full system deployment in each state.
- Annual Maintenance & Support: Ongoing bug fixes, updates, scaling.
- Customization Services: Adapting system to state-specific needs.



RESEARCH AND REFERENCES



Innovation and Uniqueness

- First AI-based occupation classification tool in India aligned with NCO-2015.
- Built for scale, capable of handling nationwide census level data without manual bottlenecks.
- Field-centric design ,optimizes for on-ground conditions (mobile use, offline capabilities, voice input for low-literacy users).
- High Accuracy + High Speed combination, rare for government AI tools (99%+ accuracy).
- Continuous improvement through user feedback, manual override tracking, and adaptive AI models that incorporate suggested terms over time.

Research Sources

- NCO-2015 Ministry of Labour & Employment, India https://labour.gov.in/nco
- Devlin, J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Google AI). https://arxiv.org/abs/1810.04805
- Singapore GovTech NLP for citizen services. https://www.tech.gov.sg
- The Economic Impact of Artificial Intelligence on Public Services - OECD

https://www.oecd.org

• McKinsey & Company - AI in Public Sector reports.

https://www.mckinsey.com



