STANDARD OPERATING PROCEDURE FOR CONDUCTING HACKATHON

Contents

Standard Operating Procedure (SOP)

Hackathon for DIID, Ministry of Statistics and Programme Implementation (MoSPI)

1. Introduction

The purpose of this document is to outline the Standard Operating Procedure (SOP) which provides a structured approach for organizing a hackathon. The manual defines key processes, roles, and responsibilities for each of the stakeholders to conduct Hackathon. The SOP should ensure a transparent, efficient, and impactful event that aligns with Hackathon objectives.

2. Hackathon Process flow:

The major stakeholder for conducting a hackathon is MoSPI and partner organisation. The chronology of conducting the hackathon is detailed below.

2.1. Pre- Hackathon activities performed by MoSPI (DIID)

a. Problem Identification and Curation

- Formulate concise, well-defined, and achievable problem statements based on the discussions.
- Preparation of relevant sample datasets for each finalized problem statement in terms of trained and test datasets.
- Define and finalize evaluation criteria such as innovation, scalability, data handling capability, and relevance to the Ministry's objectives and performance on test dataset.

b. Partner Identification and approval

- Identify the partner, who has prior experience to conducting the hackathon.
- The proposals from partner organizations should detail their proposed costing, support planning, timelines, hackathon structure and execution strategy for the hackathon.
- Solicit approval from competent authority on the received proposal.
- Release of advance payment to partner organization, if any, as per approved payment schedule.

c. Shortlisting criteria

- The screening criteria for shortlisting the applied participants must be transparent, well defined and vetted by the MoSPI. Equal chance must be provided for each problem statement and to each team/ participant.
- The evaluation criteria for selecting the winners from the final participants/teams will be done as defined evaluation criteria.

d. Interaction with the Participants

- Organize optional pre-hackathon training sessions on relevant topics such as data science tools, data visualization techniques, and APIs used within or relevant to MoSPI's datasets and platforms.
- Provide access to sample datasets or sandbox APIs in advance (where applicable), enabling participants to familiarize themselves with the data and technical environment prior to the. event

e. Mentors and Jury

- Select mentors and jury having domain knowledge for each problem statement, preferably from the ministry side.

Mentors

- The mentors need to have session with the participants to align them with the scope of the problem statement. They need to guide the participating teams on both technical and business aspects, helping them refine their prototypes into viable solutions.
- Regular interaction with the participants may be needed to resolve any issue raised by them. Mentors may provide regular feedback and checkpoints throughout the hackathon to ensure teams stay aligned with goals and timelines. Connect teams with relevant resources, such as datasets, tools, domain experts, or funding avenues, as needed for continued progress.

Jury Members

- Evaluate team pitches during the final presentation round, focusing on problem understanding, innovation, technical execution, and applicability.
- Score entries using a structured evaluation matrix aligned with pre-defined criteria.
- Provide detailed feedback and recommendations for pilot support, refinement, or further development, especially for top-performing solutions with real-world potential.
- Shortlist the most promising ideas and finalize the top solutions for recognition and potential implementation.

f. Publicity and targeted outreach

- Leverage MoSPI's official social media handles to amplify outreach and attract wide participation from students, researchers, startups, and innovators.
- Direct email outreach to academic institutions including MoU partners & institutions, and startup incubators, inviting participation and collaboration.

g. Monitoring & Reporting

- Conduct weekly planning and progress meetings with all key stakeholders to ensure alignment and timely execution of tasks.

- Organize post-event debrief sessions with internal and external stakeholders and prepare a comprehensive impact report, highlighting key outcomes, learnings, and implementation roadmaps.

2.2. Pre- Hackathon activities performed by Partner organisation

a. Proposal Submission
   - Submit a comprehensive proposal detailing project timelines, cost estimates, deliverables, and all relevant aspects.

b. Event Collateral and Platform Setup
   - Design and prepare marketing collaterals, including email templates, slide decks, and standardized templates for mentors and jury.
   - Select an appropriate platform for registration, team submissions, and judging such as the Smart India Hackathon portal, MoSPI's internal portal, or a reliable third-party solution. Alternatively, develop a dedicated microsite or landing page featuring the event timeline, problem statements, FAQs, and a registration form.

c. Marketing and Outreach
   - Develop a calendarized outreach plan, spanning at least 30 days prior to the hackathon launch, outlining targeted communication activities and engagement milestones.
   - Organize informational webinars and orientation sessions to address participant queries, clarify problem statements, and enhance the quality of participation.

d. Team Formation and Support
   - Facilitate team formation by enabling networking opportunities, interest-based matchmaking, or pre-event orientation sessions.
   - Publish detailed guides and resources, including best practices, pitch deck templates, and submission protocols to assist participants throughout the hackathon.
   - Assign a dedicated Point of Contact (PoC) or Nodal Officer to ensure smooth and timely communication between participants, mentors, and organizers.

e. Registration Process
   - Provide registrations platform for both individuals and teams, with each team comprising 2 to 5 members.
   - Capture essential details during registration, including team composition, a brief solution approach, and participant profiles (educational background, skills, etc.).

f. Screening and Shortlisting
   - Defined the screening and shortlisting criteria for all submissions and evaluate entries based on innovation, feasibility, and alignment with the problem statement.
   - Shortlist the top 15–20 teams (or as appropriate, based on the number of problem statements) for final participation in the hackathon.
   - Communicate results to selected teams, along with clear onboarding instructions, timelines, mentorship information, and event expectations.

g. Constitute organizing Committee
   - Constitute a core organizing committee comprising representatives from MoSPI and partner organisation.

- Oversee key functions such as marketing and outreach, participant engagement, , and budgeting.
- Ensure adherence to the Standard Operating Procedures (SOPs) and drive the timely and effective execution of all hackathon activities.

h. Constitute Evaluation Committee (Jury)
- Formation of evaluation committee comprising experts from academia, industry, and the Ministry.

i. Logistics coordination
- Ensure food and stay for all the participants/teams.
- Ensure accommodation and travel for outstation jury members and mentors.

3. Tentative Timeline & Milestones

S. No.	Activity	Timeline (Days)
1.	Release of advance fund as per agreement	Up to 21 days
2.	Call for Proposals / Registration Launch	Up to 7 days
3.	Last Date for Application Submission	Up to 30 days
4.	Screening and Shortlisting	Up to 7 days
5.	Orientation session to selected teams with mentors	Up to 2 days
6.	Concept note submission by Selected teams	Up to 2 Days
7.	Demo Day & Winner Announcement	Up to 5 Days
8.	Post-event Documentation and Monitoring	Up to 30-45 days

4. Tentative Budget Heads

S. No.	Tentative budget head	Estimated Budget (INR)
1.	Pre hackathon expenses	
2.	Infrastructure, Logistics (Stay), Event Hosting	
3.	Hackathon organization (welcome kit, venue preparation, AV stand etc.)	
4.	Travel-Local and outstation	
5.	Technology Platform & Online Tools	
6.	Honorarium for Experts, Mentors & Judges	
7.	Financial Awards to Winners	
8.	Post-Hackathon Pilot / Incubation Support	
9.	Miscellaneous / Contingencies	
10.	Total Estimated Budget	

Note: This is the tentative budget head and can be modified depending upon the Hackathon.

5. Hackathon Execution

5.1. Kick-off Session

- Formal inauguration with participation from senior MoSPI officials and partner organizations.
- Briefing session to align participants on rules, timelines, mentorship, evaluation criteria, and event structure.

5.2. Mentorship and Development

- Each team assigned a dedicated mentor for technical and domain-specific guidance.
- Regular monitoring through progress updates, check-ins, and interim reviews.
- A dedicated helpdesk will be available to support participants with technical and general issues.

5.3. Evaluation Process

- Final presentations by shortlisted teams to a jury panel, showcasing their prototype and approach.`
- Time-limited presentations (10 minutes) including Q&A.
- Evaluation criteria: problem understanding, innovation, technical feasibility, statistical/data methodology, and UI/UX design.

5.4. Valedictory Session

- Announcement of top 3 winning teams based on jury scores.
- Distribution of certificates and awards in a formal closing ceremony.
- Possible cash prizes or incubation vouchers to promote further development and implementation.

6. Post-Hackathon Activities

Post-event activities are essential to translate hackathon outcomes into real-world value. The following steps are critical:

6.1. Post- Hackathon activities performed by MoSPI (DIID)

a. Pilot Implementation Framework
   - Identify the top 3–5 solutions with strong potential for real-world implementation based on jury feedback and feasibility assessments.
   - Define clear success criteria, pilot timelines, and assign responsible teams for monitoring, support, and feedback throughout the implementation process.

b. Mentorship and Incubation Support
   - Provide internship opportunities to interested winners in order to helping them to further develop and implement their solutions within MoSPI.
   - Onboard incubators or accelerators to support the continued development of promising hackathon teams through mentorship, networking, and business support.

c. Others
   - A congratulatory letter to be sent to all the winning as well as participating institutions from ADG (DIID).

6.2. Post- Hackathon activities performed by Partner organisation

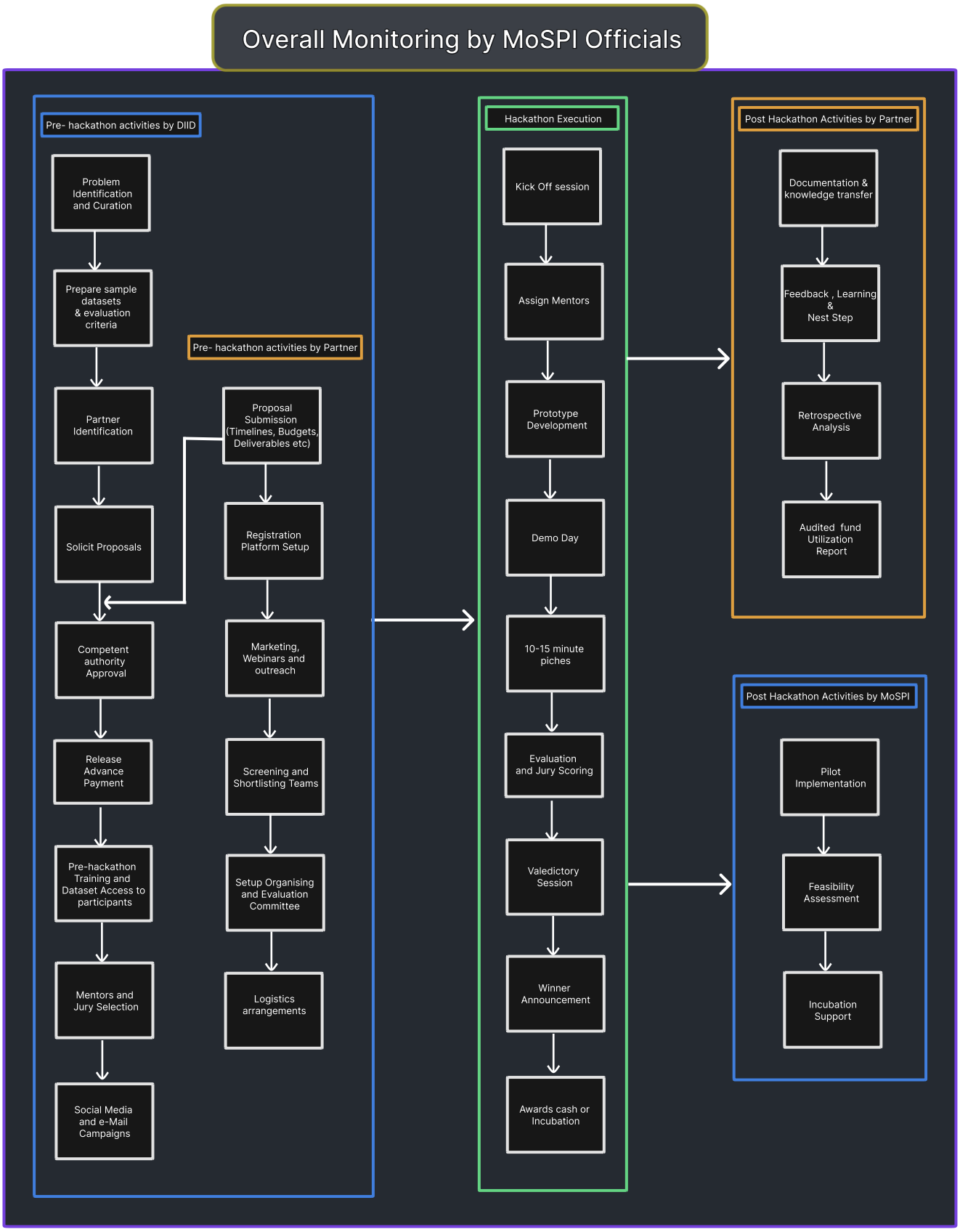
a. Documentation and Knowledge Transfer
   - Create a compendium of shortlisted ideas, including concise technical summaries of each solution include detailed architecture diagrams, screenshots, datasets used, and innovation statements that capture the uniqueness of each solution.
   - Preparation of detail hackathon report of the hackathon event along with pictures.

- Handover all documentation to the MoSPI officials for future reference, collaboration, and institutional memory.

b. Feedback, Learning, and Next Steps
   - Conduct a retrospective session with the organizing team, mentors, and key stakeholders to reflect on the planning, execution, and outcomes of the hackathon.
   - Analyze participation metrics, technical infrastructure performance, mentorship effectiveness, and evaluation workflows.

c. Others
   - Prepare a visual report and/or impact video capturing outcomes, testimonials, and success stories to be shared across MoSPI platforms and stakeholder networks.
   - Audited advance fund Utilization certification with detail item wise expenditure statement handover to MoSPI officials.
   - Complete disbursal of prize money.

7. Hackathon Flow chart



Annexure

Sample Problem Statement

AI-enabled semantic search solution for National Industrial Classification (NIC)

Details	
Particulars	**Details**
Name of Ministry/Department/Division	Ministry of Statistics and Programme Implementation/Computer Centre
Address	10, East Block, Rama Krishna Puram, New Delhi, 110066
Name of the Nodal Officer and Designation	
Phone Number (Nodal Officer)	
Email ID	
Domain/Area: Problem Statement	Official Statistics in general
Category of problem statement *(Select all that are applicable)*	☑ Data Collection ☑ Data Processing ☑ Data Quality ☑ Statistical Methodology ☐ Data Accessibility ☑ Data Integration and Interoperability ☐ Data Timeliness ☐ Data Standardization ☑ Data Utilization and Analysis ☐ Data Visualization ☐ Data Transparency ☐ Technological Infrastructure ☑ Data Management ☑ Data Extraction and Pipeline Creation ☐ Others (please specify)
What kind of support do you expect from the solution giver?	☑ Research and development of new methodology ☑ Development or modification of tools/process
What kind of support/resources will you be able to share with MoSPI?	☑ Subject Matter Experts (SMEs) ☑ Existing Datasets ☑ Technical Resources

	☐ Collaborative Networks ☑ IT Infrastructure ☑ Physical Infrastructure (space for partners to work from) ☐ Others (please specify)

Problem Statement

A. Problem statement Identified (Max 200 words) *

(Write a crisp and specific problem statement identified by the Ministry/Department/Division. Include aspects such as what the problem is, whom it impacts, and scale of impact. Try to give data figures wherever possible. Make sure the problem statement is related to official statistics areas.)

The National Industrial Classification (NIC) codes serve as a crucial statistical standard for categorizing economic activities in industrial surveys, labour statistics, national income estimates, and population censuses. These codes follow a hierarchical structure, ranging from broad sections (A to U) to granular 5-digit subclass levels. Enumerators collecting data or users entering information on portals must select the appropriate NIC code based on the given text descriptions.

With over 500 subclass-level codes, identifying the correct NIC code relies on an exact keyword-based search of text descriptions at the class or subclass level. However, economic activities often cannot be fully represented by the predefined text descriptions. Users may need to interpret synonyms or related terms, which do not always have an exact match in the NIC list. This ambiguity leads to potential misclassification, inaccurate data capture, and reduced efficiency.

The current process demands manual interpretation, consuming significant time and increasing the risk of errors. Misclassification can distort economic insights and impact policymaking. An improved solution is required to enhance searchability, minimize interpretation challenges, and ensure accurate classification of economic activities.

Contains 178 words

B. Methodology used for identifying the problem statement. *

(Detail how the problem was identified, including any studies conducted, resources referred to, or methodologies applied.)

The problem was identified from inputs from experience of enumerators and feedback available from enumerators or users and as well as from the supervisors who conducts trainings of the enumerators for NIC code Classifications. Also, whenever the NIC codes or descriptions changes, the concerned officials must remember the changes. It again takes time to memories the codes in the data collection exercise.

C. Challenges imposed and needed for solving them. *

(Explanation of the current situation, including relevant data and statistics that highlights the need for addressing this problem. List all key stakeholders affected by this problem, including internal teams, external partners, or end-users. Highlight the potential long-term impacts if the problem remains unsolved.)

Any solution available as of now for NIC code search is available with syntax-based keyword

matching search option like searching using exact keyword match in the pdf document of the list. The current process lacks semantic based search option and the enumerator/user either needs to remember the exact words to find the correct Industry code or have to search entire list and select the code. It requires the survey or/user to be experienced enough in language interpretation and trained/aware enough about the economic activities available in the NIC code list. It is a time-consuming process and may result in an inaccurate data collection. If this remains unsolved, it may lead to misleading data Insights required for policy formulation and execution and productivity loss of enumerators/users.

Key Stakeholders Impacted:

Enumerators, MoSPI and all ministries/departments, research institutions/think tanks relying on various surveys where economic activity specific Information is captured and used for economic and policy formulation and research.

D. Existing processes/systems in place to deal with the challenges (Max 150 words). *

(How is the Ministry/Department/Division currently addressing the problem statement? In case no way has been found to manage it, kindly mention that as well.)

Currently, enumerators rely on their experience to identify the closest matching sub-division within the NIC classifications. This expertise is developed over time, primarily through trial and error, as agents gradually learn to navigate the NIC list. However, there is no formal system in place to accurately track or standardize this process, leaving NIC code selection process largely dependent on individual familiarity and intuition.

. *Contains 63 words*

E. Expected outcome(s) for stakeholders' post resolution. *

(Clearly outline the benefits and improvements the impacted stakeholders will experience once the problem is resolved. Also mention the essential features of the solution.)

1) **Improved Efficiency:** An AI/ML solution with advanced keyword and context-based search will allow enumerators/users to quickly retrieve precise classifications, reducing the time spent on manual searches and enabling faster data collection.
2) **Enhanced Accuracy:** With automated search assistance, agents are less likely to misclassify entries, ensuring higher data accuracy and reliability in classification, which is critical for economic indices and policy analysis.
3) **Better Resource Utilization:** Field agents can redirect time and effort from tedious document searches/doubt clarification time to other survey tasks, increasing overall productivity and enhancing the quality of data collection.

F. How urgent do you consider it to solve this problem? *

(What degree of impact does the problem have on operations?)

☐ **High Priority:** The problem significantly impacts daily operations; needs immediate attention.

☑ **Medium Priority:** The problem affects productivity or efficiency but does not halt operations. It should be addressed within a reasonable timeframe.

☐ **Low Priority:** The problem has minimal impact on overall operations and can be resolved later without major consequences.

G. Expected timeline for resolution?

(Mention the duration within which you expect a resolution.)

A resolution is expected between 4 to 6 months.

H. Share any global best practices you'd like to highlight?

(Mention any global best practices you know of that could address your issue or be implemented to solve it. Include links where possible.)

Globally the several software companies implemented generic search engines as well a few NSO's/Statistics Bureaus implemented semantic based search for their resources in their portals or applications. E.g. Kenya National Bureau of Statistics implemented the semantic based search in their portal for searching open data available in the country. For a query "Electricity Production", The following results are returned.

The product and generation keywords are semantically related, therefore returned as the top result.

Similarly, Office of national Statistics, UK developed semantic search facility for SIC codes available in UK. Classify AI is designed to classify the SIC codes based on text input.

(Complete this section only if the Ministry/Department submitting the proposal has potential solutions in mind)

A. Proposed solutions (Max 200 Words)

(Provide an overview of the proposed solutions, including the key milestones and tentative timelines for each phase of implementation. Try to post your idea in points /diagrams /infographics /pictures.)

The solution will include an AI-based app with an intelligent search feature for conducting contextual searches. A user needs to enter/speak the text/keyword within the search bar with the closest keyword to the code description available as free text. The implementation of underlying Information retrieval system requires following phases:

1) **Data Preparation Phase:** NIC 2008 data tables available as pdf files need to be ingested into a database. Any other relevant text description data where user input as a free text was captured in the past may help in training the ai/ml algorithms.

2) **Feature Engineering Phase:** Text based features can be selected as seed words for each industry and their synonyms/related text can be generated from alternate sources of data like web or ai enabled data generation open-source tools or free text Inputs received from users during keyword-based searches for codes.

3) **Modelling Phase:** The ml-based information retrieval models like vector space models, probabilistic models, latent semantic analysis, neural ranking models etc. can be Implement to retrieve the top-k best matches for the corresponding economic activity.

4) **Model Performance Evaluation Phase:** The model will be evaluated on parameters like accuracy, precision, and recall. Based on trade-off between precision and recall, a suitable hyperparameters of the model will be selected. Further the evaluation will also consist of how well the model ranks the NIC codes i.e. its semantic search relevance.

5) **Model Deployment Phase:** Since the model will be used as an enhanced feature of already existing data collection application, it should be deployed either in the web application or mobile application.

B. Analysis of the feasibility of the solution

(Evaluate the viability of the proposed solution, considering its technical, financial, and operational aspects, along with identifying potential challenges and risks.)

The proposed solution is technically feasible leveraging open-source NLP and ML algorithms. The solution will be a standalone application with ml features from open-source models. The solution can be feasible using LLM as well as open-source solutions available for NLP like spacy etc.