CSE-400

Final Year Research & Design Project

Bangla Resume Ranker with Explainable AI(XAI)

Group Members:

Mir Sayef Ali - 202214069

Lt. Shoaib Ahmed Sami - 202114182

Ruslan Sunbeeb - 202214116

Supervised by:

Lt. Col. Muhammad Nazrul Islam

Department of CSE, MIST

Table of Contents

Introduction

Context and Motivation

- Online recruitment is increasing rapidly in Bangladesh.
- Handling large volumes of CVs manually is inefficient and slow.
- English language recruitment tools exist but Bangla tools are rare and weak.
- Lack of Bangla linguistic resources causes poor performance in automation.

Why Is It Necessary

- Existing systems rely on simple keyword matching.
- Keyword methods ignore:
  - context
  - semantic meaning and
  - different phrasing of same skill.
- This leads to:
  - missed qualified candidates
  - biases and
  - inconsistent ranking.
- Hence we need a Bangla-specific, context-aware and unbiased AI model.

Related Work

Related Work

Title	Objective	Methodology	Key Outcome (Findings and Limitations)
Resume Analysis using HDFS, MapReduce and R[1]	Estimate technical skill distribution from large resume datasets.	Preprocess resumes → store in HDFS → MapReduce counts skills → visualize in R.	Scales well for big data but only counts keywords; no semantic or contextual matching.
Comparative Analysis of CV Matching Tools[2]	Compare automatic candidate, job matching quality of three commercial tools.	Match real CVs to real job posts → compare tool rankings vs human judgment.	Better parsing = better ranking; systems still rely heavily on keyword matching.
Online Resume Parsing using Text Analytics[3]	Build a resume parser that scores candidates for quick shortlisting.	NLP extracts fields → keyword matching with weighted knowledge base → score candidates.	Automates parsing but remains keyword-based and weak in semantic understanding.
BERT: Deep Bidirectional Transformers[4]	Improve language understanding through deep contextual embeddings.	Pre-train Transformer with MLM + NSP → fine-tune on downstream tasks.	Achieves SOTA NLP performance but trained mainly on English, not Bangla.

Related Work

Title	Objective	Methodology	Key Outcome (Findings and Limitations)
Automated Candidate Selection (Bangla)[5]	Create a Bangla-specific automated CV shortlisting system.	Word2Vec embeddings + cosine similarity between CVs and job descriptions.	Works for Bangla CVs but relies on lexical similarity; lacks deep context.
Transformer Models for Long Document Matching[6]	Analyze transformer vs simple models for long-form document matching.	Compare BERT/Longformer/SMITH with GloVe/Doc2Vec models on long-text datasets.	Simple models often outperform Transformers for long texts and are more efficient.
Bangla-BERT[7]	Develop a monolingual BERT optimized for Bangla.	Build Bangla corpus → pre-train Transformer → fine-tune for Bangla NLP tasks.	Strongly outperforms multilingual models; best semantic understanding for Bangla.
Ranking-Based Bangla Summarization[8]	Improve Bangla summarization using ranking of multiple model outputs.	Generate summaries from several transformers → rank them using similarity metrics.	Ensemble ranking yields better summaries; shows ranking improves transformer outputs.
Cluster-Based Resume Ranking[9]	Improve recruitment accuracy using clustering and ML ranking.	Preprocess resumes → cluster features → compute ranking index by similarity.	Helps reduce screening workload but depends on feature engineering; limited semantics.

Related Work

Early CV Matching Approaches

- Early systems used keyword counting with fixed scores.
- MapReduce-based systems extracted skill frequencies.
- Lightweight but treated ALL keywords as equally important, no context understanding.

Related Work

ML & Statistical Approaches

- Clustering approaches attempted context-aware grouping.
- SVMs, Random Forests improved classification accuracy.
- Parsing accuracy remained critical but semantic variation still caused failure.

Related Work

Transformer & Bangla-Specific Advances

- BERT revolutionized semantic similarity tasks.
- Some cases: simpler embeddings (Doc2Vec/GloVe) outperform BERT for long documents.
- Bangla-specific progress:
  - First Bangla CV system used Word2Vec and cosine similarity.
  - Bangla-BERT (40GB monolingual model) improved performance significantly.
  - Ranking-based ensemble techniques improved summarization for Bengali.

Problem Statement

Problem Statement

Bengali CV matching is still highly inaccurate because existing systems rely mostly on keyword search, poor parsing and limited language resources, causing missed matches, bias and heavy manual screening. Unlike English, Bengali lacks robust NLP models and annotated datasets, making it difficult to understand semantic relationships between job requirements and candidate skills. Hence, there is a clear necessity for an efficient, fair and context-aware Bengali CV matching solution that can accurately analyze diverse CV formats, rank them according to the job description and provide feedback using explainable AI(XAI) to reduce recruiter and employer workload.

Research Gap

Research Gap

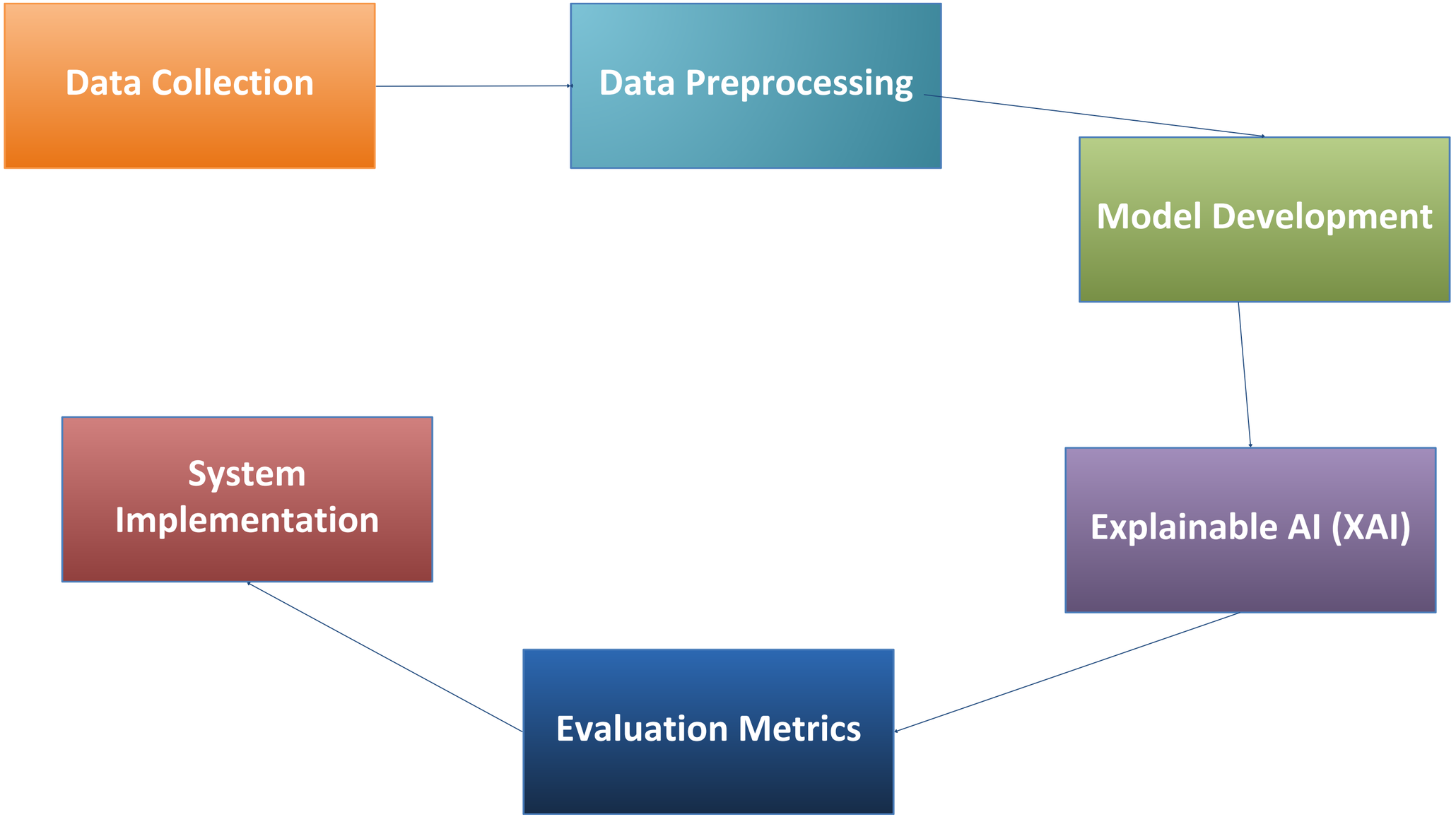
- No robust Bangla transformer-based CV matching system exists.
- Lack of:
  - annotated Bangla CV datasets
  - context-aware ranking for Bangla
  - explainable decision-making
- Existing solutions do not handle:
  - diverse CV formats
  - varied wording
  - relevance + recency evaluation

Objectives

Objectives

- To design a model that can read CVs and job descriptions with human-like understanding of skills, qualifications, and work experience.
- To automate the extraction of relevant information from CVs and rank them based on predefined criteria.
- To ensure the results are explainable so that recruiters can trust and interpret the ranking decisions.

Methodology


Data Collection
Data Preprocessing
Model Development
Explainable AI (XAI)
Evaluation Metrics
System Implementation

Timeline & Budget

Progress So Far

Expected Outcomes

Expected Outcomes

- An accurate, fully automated CV ranking system.
- Shortlisting with clear, explainable ranking criteria.
- Robust performance across diverse and unstructured CV formats.
- Adaptability to multiple job categories.
- Demonstrated improvement over traditional screening methods.
- Foundation for integration into job portals and HR management systems.

Conclusion

Conclusion

This research develops a high-accuracy, fully automated CV ranking system tailored for Bangla recruitment. The model reduces the limitations of manual screening by providing consistent shortlisting along with clear explanations for each ranking. It remains robust across diverse and unstructured CV formats and adapts effectively to different industries and job roles. The results highlight the practical value of AI-driven evaluation in local-language contexts and demonstrate improved performance over traditional methods. Overall, the work establishes a strong foundation for integrating AI-based ranking into job portals, HR platforms, and public-sector recruitment systems in Bangladesh.

References

References

[1] Krishna Mohan Ankala and Sowmya Karra. *Resume analysis for skill-set estimation using HDFS, MapReduce and R*. In Proceedings of the World Congress on Engineering and Computer Science, volume 1, 2016.

[2] Florian Buttiker, Stefan Roth, Tobias Steinacher, and Thomas Hanne. *Comparative analysis of tools for matching work-related skill profiles with CV data and other unstructured data*. University of South Florida (USF) M3 Publishing, 5(2021):97, 2021.

[3] Divyanshu Chandola, Aditya Garg, Ankit Maurya, and Amit Kushwaha. *Online resume parsing system using text analytics*. Journal of Multi-Disciplinary Engineering Technologies, 9, 2015.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *BERT: Pre-training of deep bidirectional transformers for language understanding*. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, 2019.

References

[5] Md Moinul Islam, Farzana Yasmin, Mohammad Shamsul Arefin, Zaber Al Hassan Ayon, and Rony Chowdhury Ripan. *An automated candidate selection system using Bangla Language Processing*. In International Conference on Intelligent Computing & Optimization, pages 1071–1081. Springer, 2020.

[6] Akshita Jha, Adithya Samavedhi, Vineeth Rakesh, Jaideep Chandrashekar, and Chandan K. Reddy. *Transformer-based models for long-form document matching: Challenges and empirical analysis*. arXiv preprint arXiv:2302.03765, 2023.

[7] Md Kowsher, Abdullah As Sami, Nusrat Jahan Prottasha, Mohammad Shamsul Arefin, Pranab Kumar Dhar, and Takeshi Koshiba. *Bangla-BERT: Transformer-based efficient model for transfer learning and language understanding*. IEEE Access, 10:91855–91870, 2022.

[8] G.M. Shahariar, Tonmoy Talukder, Rafin Alam Khan Sotez, and Md Tanvir Rouf Shawon. *Rank your summaries: Enhancing Bengali text summarization via ranking-based approach*. In International Conference on Big Data, IoT and Machine Learning, pages 153–167. Springer, 2023.

[9] Mayuri Verma. *Cluster-based ranking index for enhancing recruitment process using text mining and machine learning*. International Journal of Computer Applications, 157(9):23–30, 2017.