

BALASAHEB NARHARI KASHID

Email-id : kashidbalasaheb1997@gmail.com Mobile No.: +91-9790837050 LinkedIn: [/balasaheb-kashid-51950a1aa](https://www.linkedin.com/in/balasaheb-kashid-51950a1aa)

WORK EXPERIENCE

■ Data Scientist at Infoedge India Limited

(July 2021 - Present)

Individual Excellence Awardee for technical innovations and contributions to company's bottom line

- **Client Delivery Engine**
 - Generated **5% of the company's total revenue**, realised a remarkable **profit margin of 400%**
 - Developed an ensemble model of **XGBoost** and **Random Forest** to optimize telecalling recommendations
 - Analysed over **100+ features**, iteratively refining the model and strategically selecting **21 key features**
 - Developed a supplementary **Hierarchical Clustering** based user-centric **recommendation system**
- **Siamese Network Based User-Item Similarity Model**
 - Achieved **5% improvement in relevance** over the **XGBoost** and **Random Forest** ensemble model
 - Conceptualized and developed a **content-based** recommendation model inspired by **siamese networks**
 - Employed **one-hot encoded representations** for attributes characterizing users and items
 - Incorporated an **undercomplete auto-encoder architecture** for efficient dimensionality reduction
- **Node2Vec-Driven Item Embeddings**
 - Exhibited **8% superior performance** over the **word2vec embeddings** on downstream supervised task
 - Implemented a **graph-based node2vec** technique to produce item embeddings
 - Leveraged the inherent structure within the data by generating **weighted random walks** on a constructed graph
- **Word2Vec-Powered Item Embeddings**
 - Improved the **precision** of recommendations by **~15%** while aiding in **scalability**
 - Developed a **collaborative filtering** model leveraging **word2vec** techniques to produce **item embeddings**
 - Curated training corpus from users' view and search histories, generating rich embeddings that reveal nuanced relationships between entities
- **Lead Matching Engine**
 - Demonstrated **30% performance enhancement** over the legacy heuristic system
 - Developed lead matching engine utilizing **matrix factorization** based **collaborative** user-item recommendations
 - Resolved the cold start issue by incorporating **content-based** approach using **attribute-attribute co-occurrences**
 - Introduced a **semi-supervised final layer** to gauge **item relevance** for individual users

PROJECTS

- **Employee Churn Rate Prediction of a Company**
 - Achieved accuracy of **83%** using **Random Forest Classifier** algorithm
 - Employed **statistical hypothesis testing** to validate the engineered features

TECHNICAL SKILLS

■ Python, SQL, Numpy, Pandas, Sklearn, Tensorflow, Gensim, NetworkX, Dash, Shap, GitLab, Docker

COURSES AND CERTIFICATIONS

- Pattern Recognition and Machine Learning
- Non-Linear Optimisation
- Natural Language Processing
- Probability, Statistics and Stochastic Processes
- Multivariate Data Analysis
- Linear Algebra

EDUCATION

Year	Program	Institute	CGPA/%
2021	B.Tech in Aerospace Engineering	Indian Institute of Technology Madras	8.24 CGPA
2015	12 th Maharashtra State Board	Maulana Azad Junior College, Aurangabad	87.69 %
2013	10 th Central Board of Secondary Education	Pravara Central Public School, Pravaranagar	9.6 CGPA

■ Secured rank of **2929** in **JEE-Advanced 2016**