

PUNEET SINHA

Mumbai, Maharashtra, India

Email: puneetsinha@yahoo.com | Contact: +91888 883 5462

LinkedIn: <https://www.linkedin.com/in/puneetsinhads/> | Github: <https://github.com/puneetsinha>

Kaggle: <https://www.kaggle.com/puneetsinha> | Stackover Flow: <https://stackoverflow.com/users/2572745/puneet-sinha> |

HuggingFace: <https://huggingface.co/puneetsin>

PROFILE SUMMARY

Seeking an opportunity to lead AI innovation as Head of AI/ML or SVP – Agentic AI Systems in a product-led enterprise. Dynamic and innovative Data Science Leader with over 15 years of experience specializing in GenAI, Deep Learning, and ML technologies. Proven expertise in developing advanced data-driven solutions that generate actionable insights and business value. Adept at leading cross-functional teams and driving organizational growth through data science excellence. Passionate about integrating emerging Agentic AI, GenAI, and LLM architectures to solve complex business challenges.

KEY HIGHLIGHTS

- Filed patent "ADAPTIVE AI FOR PROACTIVE USER ENGAGEMENT AND RETENTION" (ref: 31727/P-3) for innovative engagement techniques.
- Conference paper accepted submitted to ICTACS-2024 will be published in SN journal in December.
- Pioneered a Transformer(attention) Model published on 'Truculent Post Analysis for Hindi Text' using BERT and LSTM in EAI.EU
- Expert in building and deploying ML/Deep Learning algorithms for various applications including OCR and NLP, Extensive experience in leading data science projects, with a focus on Trust, Safety, Content Moderation, and anomaly detection.
- Successful track record in optimizing contact centre operations using Large Language Models (LLMs), Visual Models (SORA, Invedio) , Audio Models(TTS and STT) like Whisper along with Fine tuning and prompt engineering
- Demonstrated ability in Credit Risk analytics and market segment analysis, leading to significant business improvement.
- Technical reviewer of 2 Springer publication in field of ML/AI.

TECHNICAL SKILLS

- Programming Languages and Scripting:** Python, R, Bash, Unix, TensorFlow, Pytorch, Pyspark
- Machine Learning and AI Technologies:** Regression Analysis, Classification Techniques, Support Vector Machines (SVMs), Random Forest Algorithms, Bayesian Models, Bagging and Boosting Techniques, Recommendation Systems, Natural Language Processing (NLP), Optical Character Recognition (OCR) using ML/Deep Learning Algorithms
- Deep Learning Technologies:** Object Detection, BERT (Bidirectional Encoder Representations from Transformers), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) Networks, Convolutional Neural Networks (CNN) for NLP Tasks, Sequential Modelling, OPENCV4 for image processing, Language Models, Word Embedding based Models
- Large Language/Visual Models:** LLAMA , OpenAI Audio, vedio, text models, Hugging face multiple models, Gemini, Claude, Merlin, Midjourney, Dall-e, Google imagegen, bharatGPT, Mistral
- Data Engineering and Big Data Technologies:** MYSQL, BIGQUERY, Azure Data Factory, Google Kubernetes, Cloud Spanner, Cloud SQL, Compute Engine, Cloud Natural Language API, Cloud Text-to-Speech, Cloud Speech-to-Text, Cloud Data Transfer, Web Scraping
- Frameworks and Libraries:** TensorFlow, PyTorch, Pyspark, Flask, Graphana, Celonis
- Visualization Tools:** Tableau, Graphana, Celonis, Power BI
- Web Development:** HTML, Python Flask, API Design and Development, Fast API

AREAS OF EXPERTISE

AI/ML Project Leadership | A/B Testing | Marketing Optimization | Content Moderation | Brand Evaluation | Fraud Detection | Propensity Modelling | Machine Learning | Artificial Intelligence | Deep Learning | Data Engineering | Big Data | Cloud Computing | Visualization | Web Development | Recommender Systems | Upsell/Cross-sell Models | Price Optimization | Natural Language Processing | Contact Centre Optimization | Credit Risk Analytics | Market Segment Analysis | Gen AI | Strategic Planning and Implementation | Risk Management | Operational Optimization | Team Management | LLM Finetuning | Responsible AI | AI with Moderation | Preventing Hallucinations

WORK EXPERIENCE

Associate Director- AI | Sirrus.ai, Yukio, Ziki

Apr 2024 – Present



sirrus.ai

Agentic AI & GenAI Projects :

Conversational AI Assistant in Service Commerce :

Situation: Led development of intelligent conversational AI for service commerce to enhance customer engagement and streamline bookings.

Task: Design and deploy multi-agentic conversational AI framework with state management to interpret user intent, maintain context, and integrate with business systems.

Action: Led a team of 6 AI/ML engineers delivering GenAI applications, I:

- Architected multi-agentic framework using GPT-3.5 Turbo and Gemini 2.5 Flash with API connectivity
- Developed custom LLAMA 3.1 7b model for stateful conversations
- Implemented NLU/NLG best practices with robust state management
- Evaluated multiple LLMs with hallucination mitigation systems
- Created multilingual intent classification with consistent state management
- Tested infrastructure including Amazon Bedrock and alternatives

Result: Delivered end-to-end AI assistant that enhanced customer journey, achieving 30% adoption and 10x faster booking completion. (Kindly check Huggingface repo (<https://huggingface.co/puneetsin>))

Automated Personalized Brochure Generation

Situation: Real estate developers needed efficient, high-quality customized marketing materials for different customer segments.

Task: Create system to generate personalized brochures with tailored text and images for specific developers and customers.

Action: Led a team of 6 AI/ML engineers, I:

- Implemented text generation with Mistral 7BX8 LLM for brand-aligned content
- Integrated Midjourney and Google's ImageGen for industry-standard visuals
- Developed algorithms for segment-specific personalization
- Ensured scalable quality control for generated content

Result: Successfully deployed AI product representing brand identity for three renowned developers with highly personalized marketing materials.

Trust & Safety AI for Real Estate Society Community Forums

Situation: Real estate society community forums faced significant challenges in moderating content to ensure a safe, respectful, and compliant environment. Members often posted inappropriate content, including violence, hate speech, fraud, and other violations, which eroded trust and community standards. There was an urgent need for a robust moderation system to detect and address these issues in real time.

Task: Develop an AI-powered safety and moderation system to enhance real-time detection of inappropriate content, ensure compliance with community guidelines, and maintain a safe and respectful environment for all members.

Action:

- Led a team of AI/ML engineers to design and implement the system.
- Developed **Agentic AI**, a comprehensive moderation system that checks for: **Violence, Privacy Violation, Adult Content, Animal Cruelty (image, video, or abuse), Hate Speech, Fraud, Profanity, Personal Attack, Child Safety, Spam Messages, Disrespect, Bullying, Illegal Activities, Political Manipulation, Cultural Insensitivity, Health Misinformation, Gambling, Drugs, Good Morning (spam/irrelevant content), Weapons, Religious Comments, Theft, Abuse, Offensive Language, Nudity, Mental Health (self-harm or harmful content), Self-Harm, Alcohol, Illegal Goods.**
- Integrated Generative AI and LLMs (Llama 3.2 & Gemini 2.3 Flash) for accurate detection and reasoning.
- Implemented real-time content moderation, human-in-the-loop feedback, and automated escalations for policy violations.
- Added Agentic AI as a member in the forums to actively monitor and moderate conversations in real time.

Result: Reduced inappropriate content by 45%, improved compliance with community guidelines by 50%, and enhanced overall trust and safety standards. The system significantly improved the user experience, fostering a respectful and safe environment for all members.

Responsibilities and Key Accomplishments:

Innovative Incubation of Data Science Principles:

- Successfully initiated and nurtured the application of data science methodologies within Bajaj's Personal Loan Cross-Sell team
- Directly accountable to Business Heads for managing and mentoring a dynamic team of 12 professionals
- Devised strategic approaches to amplify the existing customer base, effectively balancing portfolio risks

- Coordinated with multiple data providers, including CRIF, CIBIL, EXPERIAN, Vahan, and MSEB, ensuring comprehensive data integration

Agentic AI & GenAI Projects : Optimization of Contact Centre Operations: team size (3 ML engineers)

- Significantly enhanced lead generation by integrating Large Language Models (LLM) such as GPT-2, LLAMA, and FLAN-T5
- Pioneered in developing agent training programs and managed to handle times through innovative zero and few-shot learning techniques
- Spearheaded prompt engineering and fine-tuning of models, achieving optimal performance
- Evaluated LLMs using HELM, Rouge, and MMLU metrics and applied PEFT and LoRa methodologies to streamline model space

Boosting Personal Loan Propensity : team size (2 ML engineers)

- Elevated the likelihood of customers opting for personal loans from 0.4% to a targeted 0.8%

Strategic Market Segment Analysis and Geographic Expansion: team size (2 ML engineers)

- Conducted an in-depth analysis of market segments, identifying growth opportunities in underrepresented areas
- Set and pursued the ambitious goal of establishing 20 new potential branches to expand Bajaj's 7% market share in personal loans

Advanced Credit Risk Analytics: team size (4 ML engineers)

- Effectively controlled credit risk, maintaining key Risk KPIs such as FEMI, CBB, o+6MOB, and 3MOB
- Enhanced amount loss management and collection efficiency, contributing to a robust risk management framework

Reduction of Promotional Complaints: team size (1 ML engineers)

- Identified and addressed customer segments prone to lodging complaints with TRAI regulators, reducing promotional complaints

AVP – Data Science | Citibank

Aug 2018 – Dec 2022



Responsibilities and Key Achievements:

Deep Learning : Innovative Image Recognition Solution:

- Addressed the challenge of automating the matching of handwritten checks, including the dollar amount and signature verification
- Developed and trained Convolutional Neural Networks (CNN) and Bidirectional Recurrent Neural Networks (RNN) models to create a sophisticated in-house Optical Character Recognition system
- This system was designed to accurately interpret human handwriting on payment and deposit checks, significantly reducing manual effort and minimizing human errors

Result: Successfully achieved an automated system that accurately processes approximately 80% of payment and deposit checks

NLP : Strategic Implementation of Natural Language Processing in Banking

- Tackled the substantial financial penalties (approx. \$100MM/year) faced by banks due to regulatory non-compliance.
- Utilized Natural Language Processing (NLP) techniques to analyse text transcripts, effectively enforcing regulatory compliance
- AI/ML pipeline integrating Generative AI and LLMs for detecting profanity, misinformation, financial fraud, and fraud in digital communications
- The project involved monitoring over 40 regulations in call centres, including the Telephone Consumer Protection Act (TCPA)
- Analytical Solution: Implemented intelligent call monitoring using advanced text analytics and deep learning architectures such as BERT, Skip-gram, Transfer Learning, CBOW with n-gram approach, Bidirectional-LSTM, Logistic, and tree-based models

Result: This approach led to a remarkable saving of \$6MM monthly, the implementation of 3 new compliance policies, the elimination of manual call listening, and the removal of third-party vendors

Operational Analytics in Call Centre Management:

- Focused on enhancing efficiency by removing Level 1 (L1) support and maximizing Interactive Voice Response (IVR) containment
- Applied predictive analytics for caller intent prediction and understanding of call complexity
- Implemented capacity planning for call centres using time series analytics and topic modelling

Result: Developed an enterprise-level text analytics tool using Shiny (R), and a comprehensive Customer Journey Analytical tool, encompassing nearly all text mining capabilities

Senior Data Scientist | Yatra.com

Oct 2017 – Aug 2018



Responsibilities and Key Achievements:

Enhancement of Hotel Website Conversion:

- Spearheaded implementation of hotel discounting strategies within a defined budget, employing genetic programming for optimization
- Developed and fine-tuned a web personalization recommendation system, focusing on up-selling and cross-selling

- Utilized tree-based models for identifying key features and the Apriori algorithm for mining sequences with high confidence and support.
- Conducted comprehensive A/B testing on various propensity models

Result: Boosted the hotel website's monthly conversion rate from 6% to 10%, generating approximately 20-25 lakh in additional revenue. Also achieved a 6% conversion rate in the churned customer segment

Optimization of Upselling and Cross-Selling in Cab Services:

- Led the analysis and utilization of tree-based models to identify actionable features for the cab business
- Implemented similarity-based approaches for identifying and targeting new customer segments with specialized campaigns

Result: Successfully increased upselling and cross-selling by 18% in a quarter, contributing to an additional 10-15 lakh in revenue per month.

Strategic Use of Time Series Analysis in Flight Fare Management:

- Analysed flight fare variations over time to create a sense of urgency for online ticket bookings
- Executed customer profiling to identify and capitalize on market opportunities

Lead Data Scientist | Infosys Ltd

Nov 2016 – Oct 2017



Key Responsibilities and Achievements:

BPOST Mail Classification System:

- Developed an end-to-end solution for categorizing BPOST's mail, including digitally scanned items

Result: Enhanced mail sorting efficiency and accuracy

ExxonMobil Fraud Detection System:

- Implemented a credit card fraud detection system for ExxonMobil
- **Result:** Improved fraud detection, reducing financial risk

Data Scientist | IBM Research

Nov 2015 – Nov 2016



Key Responsibilities and Achievements:

- Led the development and implementation of machine learning models in retail and telecom sectors
- Proficient in big data management and analysis using Apache Spark and Python tools, enhanced data processing and insight generation
- Utilized key algorithms for regression, clustering, time series, and classification to improve pattern recognition and strategic data utilization

Assistant Manager | Vodafone

Sep 2014 – Nov 2015



- Crafted analytical solutions for customer targeting, market segmentation, and product design at Vodafone Germany and Italy, enhancing marketing and product strategies
- Conducted demand forecasting and brand valuation, improving market trend prediction and brand value assessment
- Executed profit and price analyses, leading to more effective financial planning and pricing strategies
- Developed fraud detection methods in the telecom sector, notably reducing fraud incidents and enhancing customer trust

Senior Subject Matter Expert | Amdocs

Dec 2010 – Aug 2014



- Served as Senior Product SME at Amdocs, handling end-to-end aspects of telecom rating and billing products (Ensemble and Enabler), enhancing operational efficiency
- Supported IBM's statistical team with data cleaning and transformation, ensuring data accuracy and usability
- Developed data pipelines for the analytics team, streamlining data processing and analysis workflows

EDUCATION

- Pursuing **Master of Technology in Artificial Intelligence and Machine Learning** from Birla Institute of Technology and Science, India, **2025(perusing)**
- **Bachelor of Engineering in Computer Science** from CKPCET, South Gujarat University, India, **2010**

CERTIFICATIONS AND OTHER COURSES

- Advanced Certification in Generative AI with Large Language Models
- Introduction to Attention Models
- Comprehensive Course on Product A/B Testing with an Interview Guide
- Detailed Study in Time Series Analysis and Forecasting using Python
- Extensive Training in Machine Learning for Data Analysis: Focusing on Regression and Forecasting
- MLOps Fundamentals: Mastering CI/CD/CT Pipelines of ML with Azure
- Microsoft's DAT203x Data Science and Machine Learning Essentials
- Machine Learning Specialization by Coursera
- In-depth Learning in Getting and Cleaning Data
- The Data Scientist's Toolbox
- Practical Machine Learning
- Neural Networks and Deep Learning
- R Programming