Luein Analytics

Turn your data into EFFICIENCY and COMPETITIVE EDGE

AI Opportunity Analysis

AI Strategy Consulting

Custom AI Solution

" The best AI solutions are not the ones that are most talked about, it's the ones that solve YOUR problem with adherence to cost and infrastructure constraints "

Luein analytics

With deep expertise in AI research and development consulting, we help businesses to solve their most important problem statements.

About us

Luein Analytics established in 2017. A company based at Bangalore-India(HQ) and Malaysia, with its clients across South East Asia, UK, UAE and USA (Silicon Valley).

AI Consulting focused on breakthroughs in the field of Natural Language Processing, Natural Language generation, Audio & Video analytics, Image processing, and Statistics.
We are a strong team of some of the brightest minds in Statistics, Math - AI Architects, AI Strategists, NLP experts, data scientists, Computer Vision & Audio/Video experts and domain data SMEs.
Successfully delivered 100+ AI strategy and research solutions till date.

Our AI consulting services

AI Opportunity Analysis

We find you the lowest hanging fruits for AI-driven automation with clarity on the potential ROI. We also help you prepare for future automation opportunities if you're just getting started.

AI Strategy Consulting

Have an idea but don't know how to proceed? We'll bring your AI vision to life. We work with you at the planning as well as implementation level to move your AI agenda forward.

Custom AI Solution

Get complex AI systems developed by the experts. We start with your business goals, followed by an analysis of your current technology stack, data availability, and more.
Prototype | Minimum Viable Product | Complete Solution | Help with Existing Product

ISO

ISO 9001:2015 IAF-Certified
Quality Management System

ISO 27001:2013 IAF-Certified
Information Security Management System
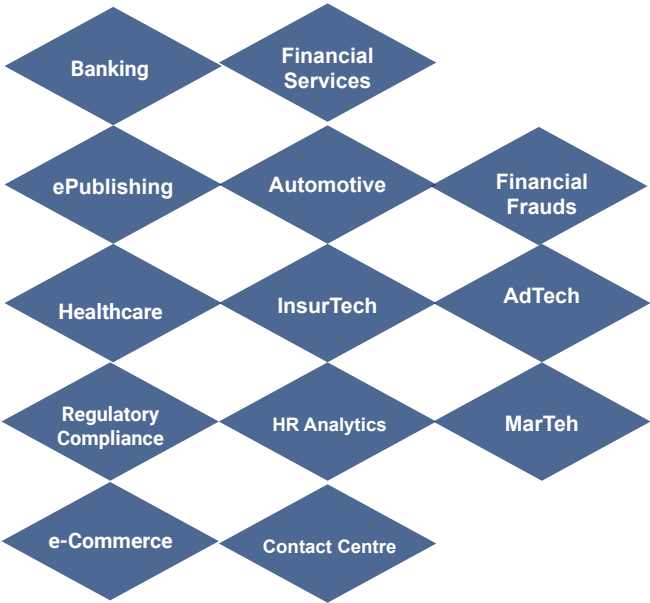
45+ AI Consultants   **50+** Clientele   **150+** Innovative solutions delivered

50+ Clientele

Domain footprints

- Banking
- Financial Services
- ePublishing
- Automotive
- Financial Frauds
- Healthcare
- InsurTech
- AdTech
- Regulatory Compliance
- HR Analytics
- MarTeh
- e-Commerce
- Contact Centre

Our AI Expertises

Standard Text Analytics

Concept Mining*
Intent detection
Language tone detection*
Grammar evaluation
Emotion detection
Sentiment analysis*
Anomaly pattern detection
Intent detection & classification
Conversational analytics*
Custom entity extraction*
Named entity extraction*
Synonyms & antonyms extraction*
Content clustering/categorisation*
Content classification*
Primary predicate detection
Languaging modeling*

NLP & NLU

Language parsing*
Grammar correction
Dependency parsing
Correlation detection
Entity linking
Topic generation*
Abstractive summarisation*
Language assessment
Content recommendation
Ontology development
Complex anomaly detection
Chat intelligence*
Plagiarism check
Market intelligence - technology scouting
Customer churn prediction

Deep Learning based NLP

Content relevancy check
Contextual similarity*
Knowledge graph
Intent search
Document similarity check*
Semantic Search*
Contextual gap estimation with semantic similarity
Content encoding*
Natural language generation
Semantically similar content augmentation
Complex languaging modeling
Conversational AI*
Auto complete & auto correction
Language translation*
Question-answer generation

Audio, Video & Image

Object detection & tracking
Image matching
Face authentication
Emotion detection OCR
Information extraction
Speaker diarization*
Audio Intent detection
Tonality detection
Confidence evaluation
Speech authentication
Emotion detection*
Multi-language processing
Median frame energy*
Mean/median/standard deviation from frame
Audio Similarity Check*
Audio Segmentation*
Audio Search

* **Multilingual Solutions**

Our AI Case Studies

- Semantic Search Engine - Regulatory Compliance
- Concept Extraction
- Chats, emails, and call transcript based customer churn prediction
- Rejected Article recommendations
- Sentiment analysis and customer satisfaction
- Journal content summarization
- Language Assessment tool
- Multilingual Transfer Desk Agent
- Email classification services
- Automated call and email analysis for Investigations
- Contextual copy-editing (grammarly alternative)

- Affiliation & Reference Structuring
- Automated call auditing to enhance customer experience for a contact center
- Technology scouting - recommending upcoming booming technologies
- Finance risk prediction from external web sources and internal transaction data
- In-hospital claim prediction
- Insurance claims prediction for car insurance firm
- Employee retention and salary prediction by analysing archived resume dataset, market trend and candidate social media behaviour
- Ancestry data mapping - name entity recognition, face detection and auto mapping

- Banking loan fraud detection from customer call behaviors
- Automatic zoning searchable PDF
- Auto Proofreading
- Invoice content extraction - scanned Images
- Smart OCR (Computer vision + NLP)
- Alt-Text generation
- Ad Extraction (NLP+OCR+web)
- Healthcare customer leakage risk prediction from agent-customer audio recordings
- Generate cross sell opportunities for a leading UAE based insurance firm
- Auto-dubbing 14000+ hours of tv series in multiple Asian languages (under development)

and more….

Semantic Search Engine for Risk and Compliance Market

Advanced AI based semantic search engine for the Risk and Compliance Market.

- Search across regulations, legislation and industry standards to identify similar requirements.
- AI powered search uses high-dimensional vectors and graph technology to fetch rules that are semantically similar to your search phrase.
- Scan authoritative sources for changes and new rules. AI powered semantic search fetches rule changes and Natural Language Processing highlights the rule text that has changed.
- Similarity scores between high-dimensional vector representation of rules, obligation statements, policies and controls powers AI to determine the exact impact of rule changes across your business locations
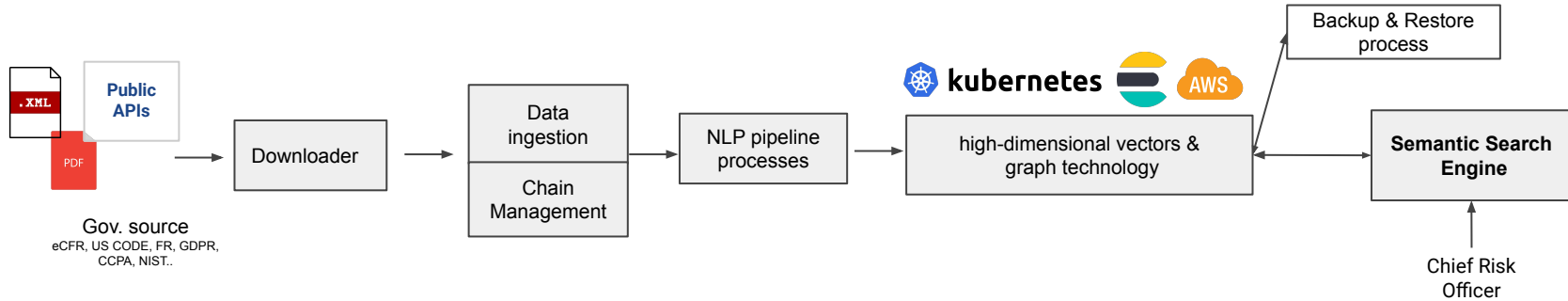
- Ensure adequate rule coverage and consistency from group to business unit policies.
- Identify weakness in your policies and controls against compliance rules.
- Generate rule summary from related requirements across regulator's law and standards.

Crop diseases detection using mobile application

Automate crop disease detection for vegetable crops
6 categories of crops data collection, AI model training and a mobile friendly, cloud based PWA (progressive web app) application developed in 3 months time, and around 1500 man hours effort. Every additional category data collection and followed AI model training/retraining took around 1.5 man weeks time. With a support of 20+ Indian regional languages and future scope of live video conferencing architecture.

1 Click photo

2 Upload for analysis

3 Receive detailed result
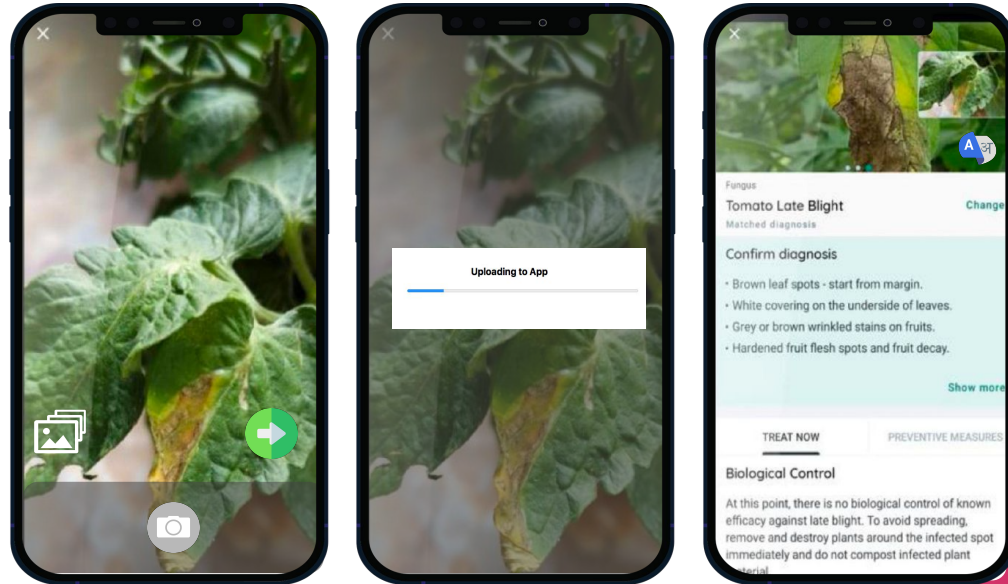
AI for Waste management - Smart Sorting

Using Artificial Intelligence and Robotics, we are building automate the solid waste sorting task, where the tool can separate recyclable commodities and non-recyclable from the solid waste endlessly. Using this duo technological enhancements, the client is trying to solve the problem fraction by fraction every coming year.

Robotics arms **#1**   AI software   Robotics arms **#2**
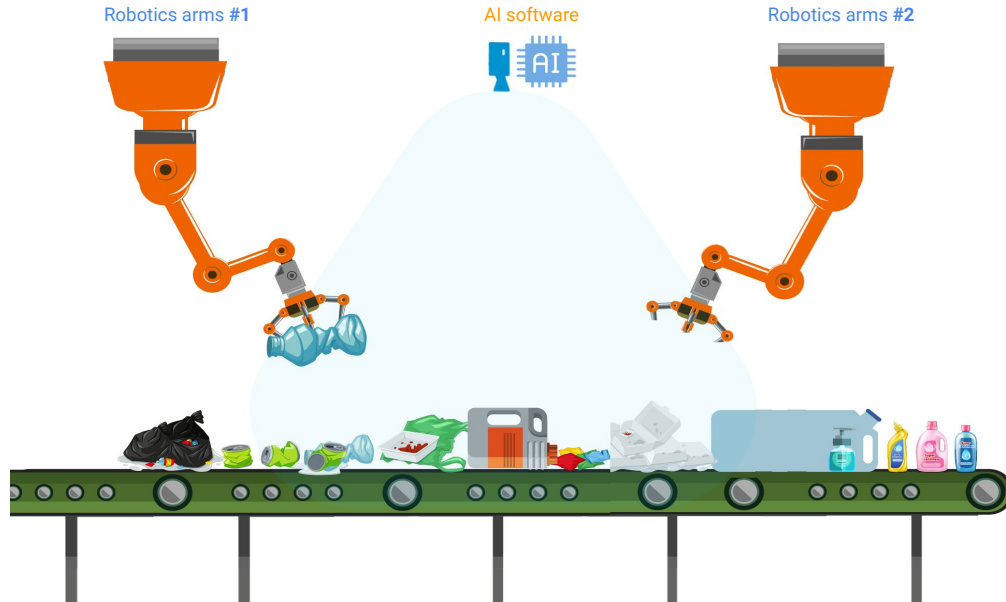
1 **Search** with AI visions

2 **Grab** with Robotics Arms

3 **Separate** the commodities for recycling

Face recognition and Geo-location based School Attendance System

Cutting edge Deep Learning based face recognition and geo-location based school attendance management system. This solution was also tested for Anganwadi in MP and Chatravas centers.

Web RESTful API (Application Programming Interface) micro-services (sync and async) Installable mobile and desktop responsive application State-of-the-art light weight deep learning face recognition algorithm Dockerized container on Linux operating system server Queueing mechanism to handle millions of simultaneous attendance requests Works on any device or external camera connected to internet - Any mobile / tablet / laptop / kiosk or external device camera with 10MP (or more) for effective long distance face recognition
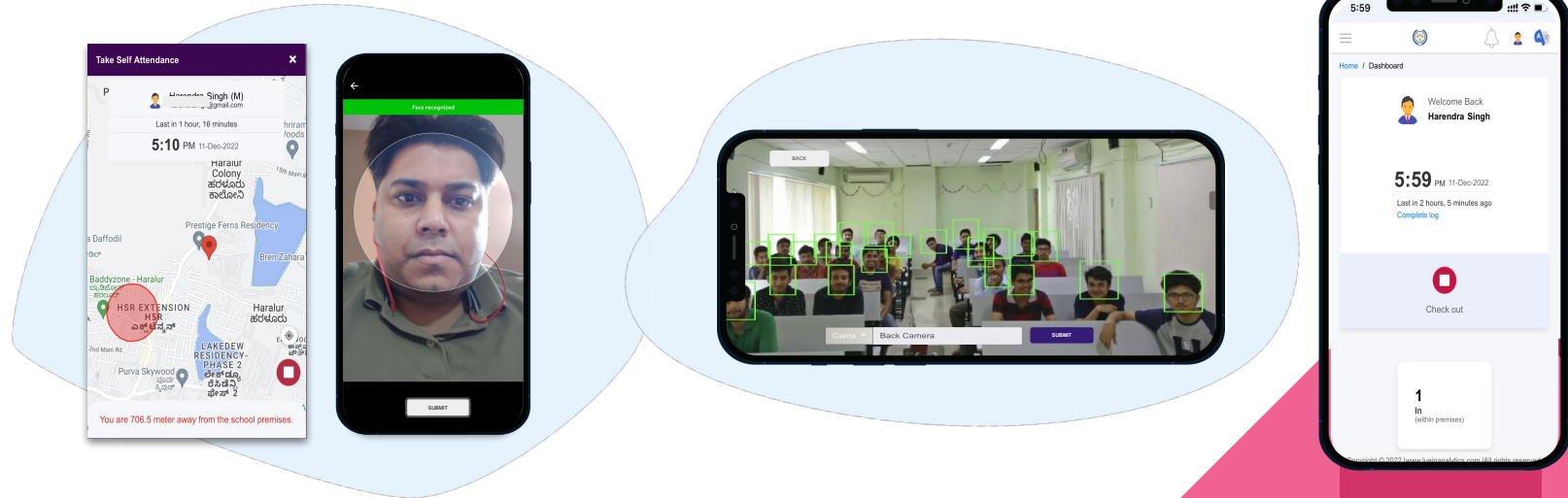
Traceability, Impact Assessment & Compliance Map

Critical challenge for a compliance team is to manually trace all internal business unit rules/policies/procedures/controls, for each and every new compliance requirement. Here is where AI powered semantic search across regulatory landscapes can trace internal policies, controls, processes, risks, etc and can map a regulatory relationship with the new external compliance requirements.

- This helps the compliance team to identify weakness in their policies and controls against compliance rules. And also ensures an adequate rule coverage and consistency across all relevant internal regulatory business units.
- The weaknesses are identified using a verified machine learning algorithm, where correlations are identified between the internal and external rules, and the correlation score is then interpreted into a relationship score/weight. Lower the score, weaker the external compliance mapping to the specific rules/policies/procedures/controls.
- The compliance teams can perform such traceability, create regulatory maps between rules & internal rules, and identify weakness in your policies and controls against compliance rules, and finally bookmark all related rules across regulator's law and standards under common tags. Such related obligation rules from regulators and from internal governance requirements can be used later for Obligation Generation.
- The background ETL change management flow will track all internal rule changes, and store the changes in a timeline manner against each selected internal rule.
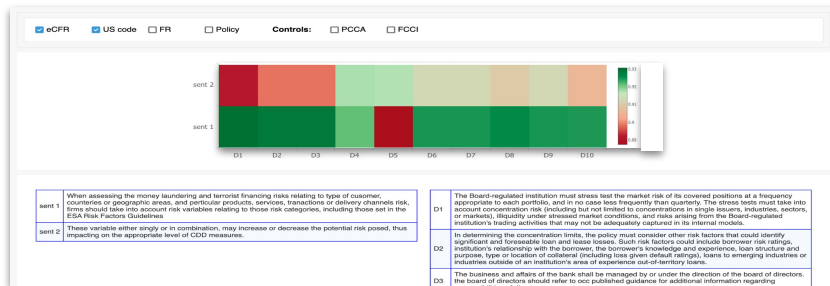
Here in the example image, the input "Sent 1" requirements show a high similarity map against the available top matching internal rules. Whereas "Sent 2" ideally is a new requirement that has to be adapted internally to meet the compliance, as the existing top contextual matching rules have a huge gap.

Complex Obligation Generation

The related obligation rules from regulators and from internal governance requirements are summarized across specific category/part/topic level to generate a business obligation.

- The generated rule summary from the bookmarked related-rules, across regulator's law and standards, can then be shared across other compliance bodies for peer review and correction.
- AI algorithms supported with business rules, helps in summarizing the obligation rules into a new obligation which has high traceability across all selected obligations and with no redundancy of the information. Keeping the main intent and key concepts from all selected obligations and then creating a new obligation is a quite complex process.
- An ensemble machine learning approach is used for obligation summarization, which start with - concept extraction, duplicate removal of concept & contexts using encoded representation of each rules, chronology and context based content binding (extractive approach), and apply a generative based business language rewriting of the final summary to generate multiple variations of it. Finally the generated obligations are then compared for traceability/overlap across all input obligations to select the one obligation with highest and the best contextually meaningfulness.

Rules	Domain	Agency	Law/Act	Product Class	Theme/Sub-Theme
Any transaction not specifically permitted in a special purpose acc Title 12 Part 1022 (Section 1022.43) https://demo1.phigrc.dev/ecl	Banking	CFPB	CFR Title..	Lending	CDD timing requirement Add/Change
Direct dispute notice contents. a dispute notice must include: this Title 12 Part 1022 (Section 1022.43) https://demo1.phigrc.dev/ecl	Banking	CFPB	CFR Title..	Insurance	CDD timing requirement Add/Change
Delivery against payment, payment against delivery, or a c.o.d. tra Title 12 Part 1022 (Section 1022.43) https://demo1.phigrc.dev/ecl	Banking	OCC	CFR Title..	Lending-Insurance	CDD timing requirement Add/Change
Margin means the amount of margin which a creditor would requ Title 12 Part 1022 (Section 1022.43) https://demo1.phigrc.dev/ecl	Banking	FRB	CFR Title..	Lending-Credit	CDD customer identity Add/Change
Buying or carrying any part of an investment contract security whi Title 12 Part 1022 (Section 1022.43) https://demo1.phigrc.dev/ecl	Banking	NCUA	CFR Title..	Investment	CDD customer identity Add/Change
"signed" means that the written attestation, declaration, or permi CCPA 1.0.999_301|u (CCPA|1.0.999_301|u)	Privacy	State of California	CCPA 301	Deposit-Now	CDD customer identity Add/Change

Category/topic/theme wise generated obligation response. Confidence score highlights the machine level prediction confidence in generating the best summary, including the intent and context of the selected rules.

Thems/Sub-Theme	Description	Confidence Score	Source	Related Rules	Jurisdiction
CDD timing requirement	Customer Due Diligence must be performed for occassional transactions which exceeded the following	95%	eCFR, GDPR	Title 12 Part 1022 (Section 1022.43) Title 12 Part 1022 (Section 1024.3) Title 12 Part 1022 (Section 1022.35)	USA
CDD customer identity	The bank is under the obligation to identify and verify the customer's identity on the basis of documents or	85%	eCFR	Title 12 Part 1022 (Section 1022.43) Title 12 Part 1022 (Section 1024.3) Title 12 Part 1022 (Section 1022.35)	USA

Automotive AI initiative for an Indonesian Motor-Insurance company

We have worked with an Indonesian motor insurance company in identifying and solving various cost saving and new innovative cutting edge AI solutions. We have covered all the stages in a claim, underwriting, and repair life cycle.

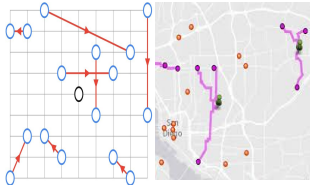KYC verification Verify Indonesian and Indian KYC documentations.
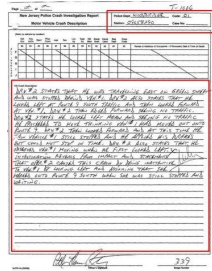
Digitize scanned insurance documents

Digitize scanned insurance documents

Car orientation detection Use the orientation to dynamically guide user to capture car 360 degree video

Home inspection Routing Problem with Time Windows Customer requests for an Inspection person to take the photo.

Home inspection Routing Problem with Time Windows Customer requests for an Inspection person to take the photo.

Concept extraction and tagging

Concept tagging and indexing process in Biomedical and Chemical research contents are very critical, tedious and long time taking task. Many organisations hires hundreds of subject matter experts specialised in these fields, to identify ideal and relevant tags for each contents.
All the concept extraction and indexing of **biomedical** & **chemical** research journal and articles were done manually, till last year(2016). Overall 150+ SMEs were involved within the selected academic areas, overall there were 1200+ SMEs involved across 40+ academic areas.

Challenge:
Research journals and articles contained quite complex literature.
Subject matter experts are familiar with the document structure, so they are experienced in extracting multiple concepts from a 2 page journal to 500 page books.

PDF

Source
Research
PDF

Pre-processing → Doc Vectorization → **ML Model** ➕ **Proximity Model** → Key Shortlisting → Operator Response Interface

Save Weights & Model
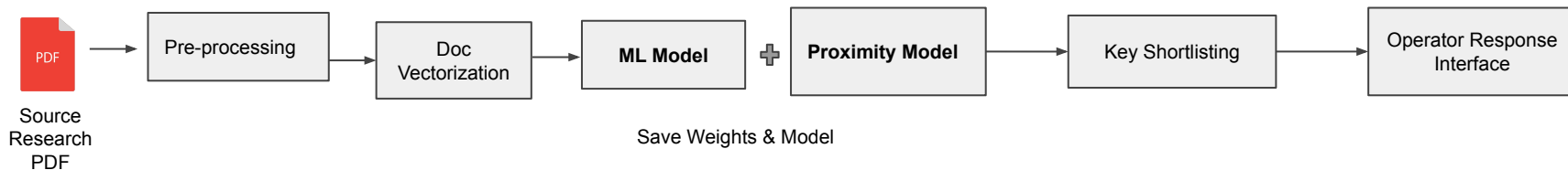
Approach:
As a starting point, we have used the readymade data sources - **biomedical** & **chemical** existing ontology dictionaries. With NLP we have increase the verified concepts in the ontology for **auto indexing** process.
Document clustering, content localisation, heuristics based concept ranking, filtration of low probability concepts and manual quality check were the steps we followed to create our own auto indexing tool.

Result
Which recommends 15-20 concepts per document, with a 85% od average selection confidence.

Chats, emails, and call transcript based customer churn prediction

Early detection of frequently discussed topics and problem in chats and emails, such as product bugs or weak spots on the website, sometimes few intimate information about their agent-customer interactions are vital to preventing customer unhappiness and reduce the customer churns.

Approach

- Data sources - Chats, emails (directed to a single company group email), call recordings and feedback forms.
- Natural Language Processing with standard text analytics techniques were used, remove noise, extract important topics and problems, detect sentiment around any discussed topic, classify discussion based on severity from all input sources.
- Computed correlation b/w the extracted topics and the existing customer churns to build a good metrix and ranking algorithm.
- Different period insights are used for to determine the long-term & short-term factors.

Result

- Gained significant speed in email and chat processing, structuring and detection critical fields (topics and problems).
- Helped the client fix the problems quickly and keep the customer interested by identifying hidden topics and problems.
- Client used the information to run a successful retention campaign.

In-hospital claim prediction

Claim predictor extract coded data for care episode, applied predictive analytics & NLP models, and provide recommendations of missing expenses/charges (like missed procedures, devices/implants, diagnostics, etc.).

Approach

1. **Extract coded data**
2. **Prediction reviews:** Coder/biller reviews and accepts the predicted missing charges if appropriate documentation is available, ensuring more complete claim on initial submission.
3. **Submission & remittance:** Revenue integrity owner utilizes the Claim Predictor analytics to assess revenue leakage mitigation and drive documentation, coding, and technology improvements.

Expense sources

Patient

Claim predictor

Claim submission

Biller/Coder
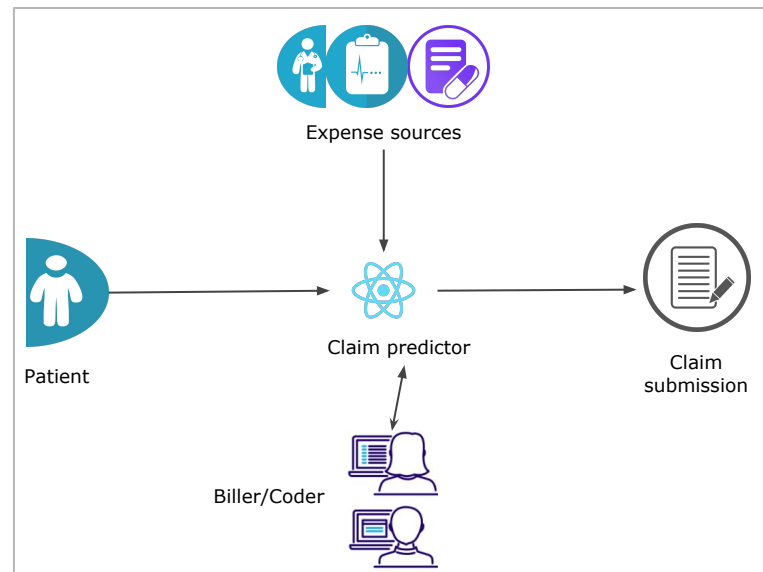
Insurance claims prediction for car insurance firm (India)

Client receives car accident insurance claim request and along with the car damaged area pictures under the claim process. Based on the extent of damage examined in the pictures, the operator decides the applicable claim amount to be give to the customer. Automate the whole manual examination process using, neural network based image processing, business rules and predictive analytics, predict the applicable allowed claim for any new request.

Approach

1. Process each incoming images, identify the different car body part and damage area for the given car images. Identify VIN, odometer and license plate details using computer vision, rule-based and NLP logic.
2. Using CNN based neural network algorithm, identify the extent of damage, and map with the business rules for the applicable claims on the same.
3. Also predict the applicable claim amount possible, on considering other car details, like manufacturing date, model, location, insurance plan, etc.
4. Predict final allowed prediction amount to the operator to initiate the claim process.

Employee retention and salary prediction

Analysing archived resume dataset, market trend and candidate social media behaviour

We built a system to tell how much a person earns based on the traces they leave on the Internet(Facebook, Instagram, Twitter and linkedin).

Approach

- Data collection - Collected 500+ employees salary, employee details from 17 companies (including startup, SME & MNC)
- We requested social media data from these 500+ employees, all participants were fully informed of the data they would be sharing when consenting to participate.
- We had gender, income, and limited personal detailed of each users.
- Analysed each individual's Facebook + Instagram + linkedIn + internal non-compliant company chats.
- Stats collected: location zip code, age, gender, ethnicity, educational attainment, occupation industry, and personality, social media platform insights.
- Based on above data, multiple models were developed - likes/dislike prediction, intent from status, to predict correlations between predicted and actual income values.

Social insight's correlation with actual income

Category	Value
Personality	0.14
Demographics	0.16
Postcode Income	0.21
Industry	0.23
Education	0.3
All Socio-Dem.	0.42
All Socio-Dem. + Facebook	0.48
All Socio-Dem. + Personality + Facebook	0.49
Likes + Status Updates	0.43
Status Updates	0.41
Likes	0.27

High income group
frequent words

Low income group
frequent words

AI powered video hiring platform (NLP + audio processing + video processing)

In the world of biased candidate assessment, influenced by a human with an unclear definition of job success, asking inconsistent questions and evaluating on unknown criteria. It is really important to involve data-driven methodologies for candidate evaluation during interview process.

Approach

1. AI based video pre-assessment solution - hire the best talent, faster, even remotely.
2. Enable video based pre-hire assessment, video interview platform and mock tele call based assessment.
3. Reduce bias and human error in the hiring process. A data-driven method that's fairer, consistent, auditable, improvable, and inclusive.
4. Opening up more opportunities for a wider variety of well qualified people.
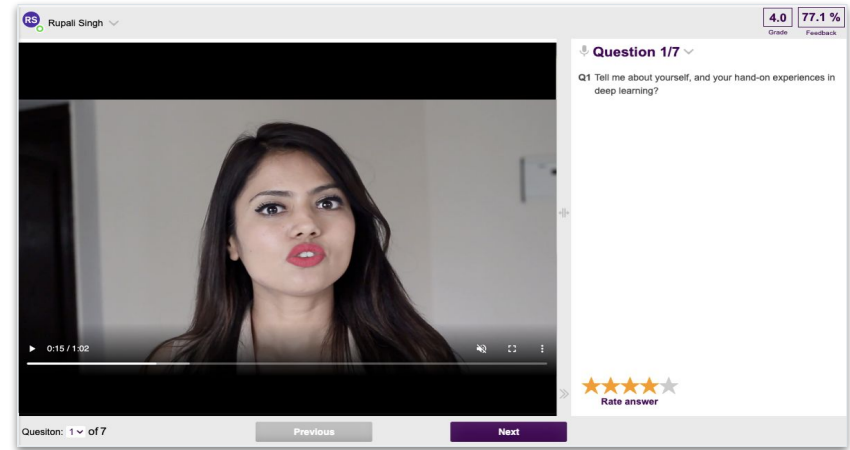
Pre-assessment

Remote interview

On demand interviews

Mock video interview

AI powered candidate psychological assessment

NLP, Video and audio analytics powered applicant and candidates soft skill psychological assessment.

Feature: Personality check , Communication check, Confidence check, Competency evaluation, On-demand response assessment and Interview intuition

Proxy interview detection - using audio-video inputs

Proxy interviews are a big pain and leads to enormous business loss directly and indirectly.

At Luein Analytics, we have tried to solving the proxy interview issue through our multilingual advanced video hiring solution.

Candidate's 2500+ facial feature movements, verbal and non-verbal actions, voice & lip activity detection, along with the spoken words are monitored at every point of time by the application, to detect any possibility of proxy interview. The system alerts/notifies the interviewer of any proxy possibilities during a two-way interview and highlighting the proxy timestamps during an on-demand one-way interview with machine.

The solution pilot run was successful for a Malaysian recruitment & staffing for their remote interview and hiring process.

{"score": 40, "prediction": "Fake", "confidence": "93.67%"}

Lip Activity Detection: (Change in Lip movement)

Voice Activity Detection: (Voice or unvoiced)
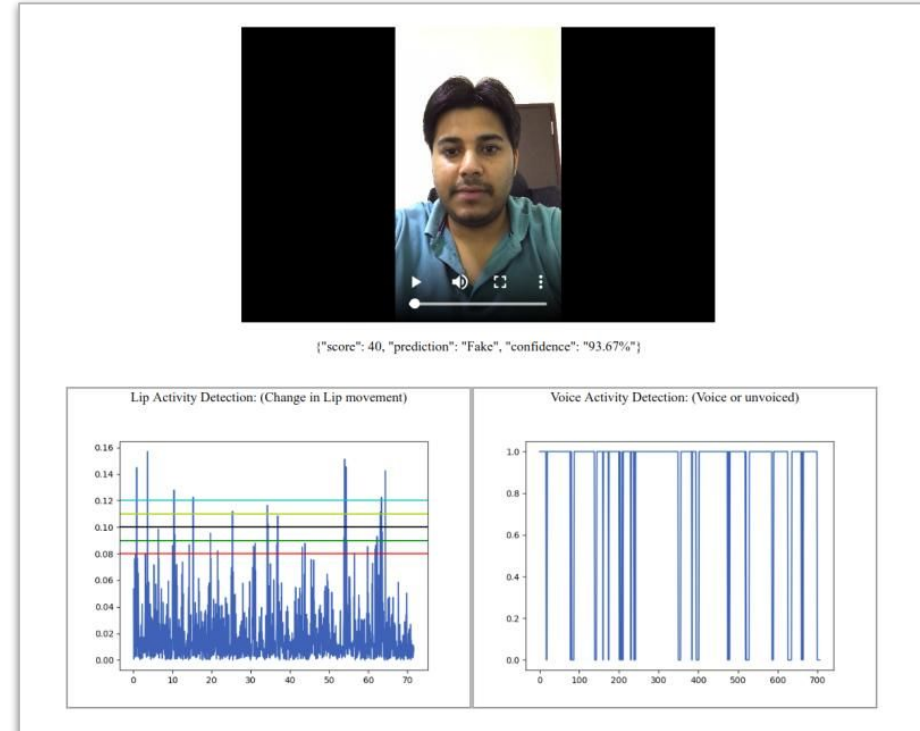
Optimized logistic packing algorithm for an Indian online grocery e-commerce

Select the best and most efficient box for a shipment

Dependent factors - shipper types, dimensions of items, rotations, packing more than one items at a time, shipping costs, operation time, and experience for both our clients and their customers.

Approach

Built ensemble models, combining below algorithms (with customization):

- **First Fit Descending**, pack the biggest products first in the smallest space we can.
- **Knapsack problem**, given a set of items, each with a weight and a value, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible.
- **Bin packing problem,** objects of different volumes must be packed into a finite number of bins or containers each of volume V in a way that minimizes the number of bins used.

Reduced Average cost per order by ~7.5% and increased Percentage of orders delivered in full by 12%; Percentage of on-time deliveries by 12% (relative)

Optimized packaging algorithm
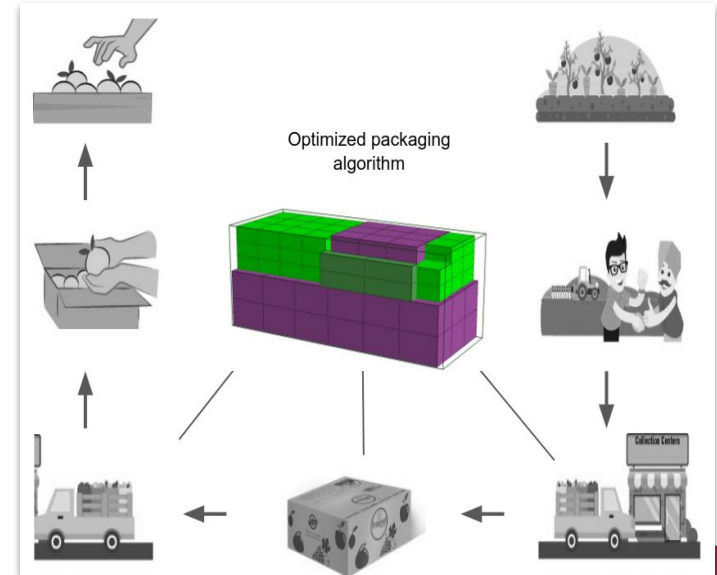
Ancestry data mapping - Named Entity recognition, face detection and auto mapping (~1.2B faces)

1. Extract faces and names from school yearbook scanned image dataset (collected from US and 2 more European countries school year books since past 45 years) and create facial database for the client automatically.

2. Detect face from individual face images and also from group scanned yearbook image files.

3. Extract Names and alias names from the same scanned yearbook image file.

4. Using natural language processing and custom business rule to associate faces with their names and feed it into the database.

Sample input image file
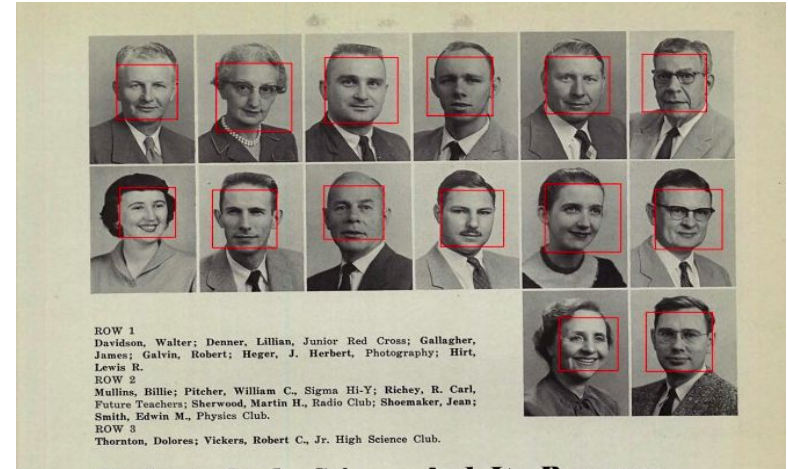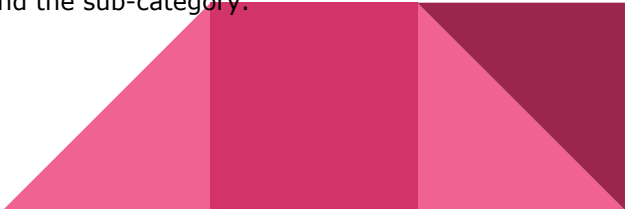
Rejected Article Recommendation

Wiley publication (digital content publisher) has multiple Journals in it, they publish multiple of articles in each journals every year. Each Journals can have many categories and sub-categories in it.

If Wiley rejects an author's article in one Journal, then this article should be checked with other Journals for any possibility in some other category before rejecting it completely from Wiley.

Approach

1. Collect all rejected journal from Wiley, with their rejected journal name and category.
2. Keep a dataset of all published articles in each Journals with its category/sub-category name.
3. Create a word embedding and 300+ dimensional classifier feature to differentiate each category and sub-categories.
4. For any new rejected article, match the document closeness and similarity with the trained category matrix, and find out which category/sub-category has more changes for the rejected article to get selected within Wiley.
5. Recommend top 5 Journal for the rejected article, along with the category and the sub-category.

Sentiment analysis and customer satisfaction

A grocery delivery company wanted to optimize the customer and delivery person interaction using their existing mobile app.

Challenge

- All delivery agents are not proficient in english language, so the way they write replies to any urgent customer query will severely impact the customer sentiment.
- Short chat message recommendation during conversation, works best for both delivery agent and customer.

Approach

- Analysed archived data (1.5 years of interaction) - 'customer - delivery agent chat conversation' and 'customer feedbacks'.
- Three models are developed for - customer sentiment detection, main intent detection (who needs what) and chat recommendation model based on 'sentiment + message context +predicted intent)
- For every critical sentiment messages, few short chat messages are generated and recommended.

Hey, I'm waiting near the starbucks

Okay, got it!   I'm on my way

Send a message...
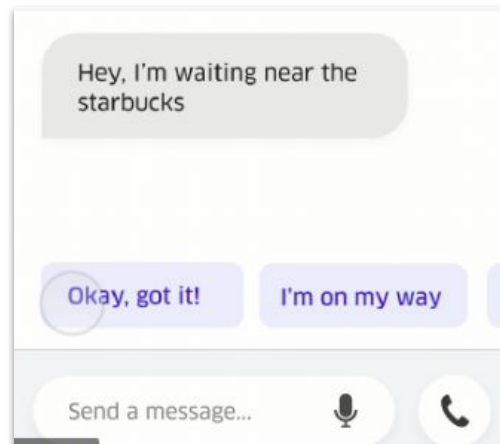
Reference : Uber & LinkedIn one click chat

Journal content summarization

We have worked on multiple content summarisation use cases till date, and there are always plenty of point of improvement and technology advancement we get when we work again on any of the summarisation use cases.

Few of the case studies are:

1. Generate initial article draft for an author, from a given topic and related key phrases

Create unique content from scratch, simulating a human writer. You choose the topic and length, and the algorithm will create your textual content.

1. Natural Language Generation for Financial Services

Automatically generate high quality personalised risk analysis, financial, compliance and other reports in writing in seconds from a given report stats, and keyphrases.

1. Academic books summarization

Automatically generate titles, rewrite articles and content summarisation for academic books.
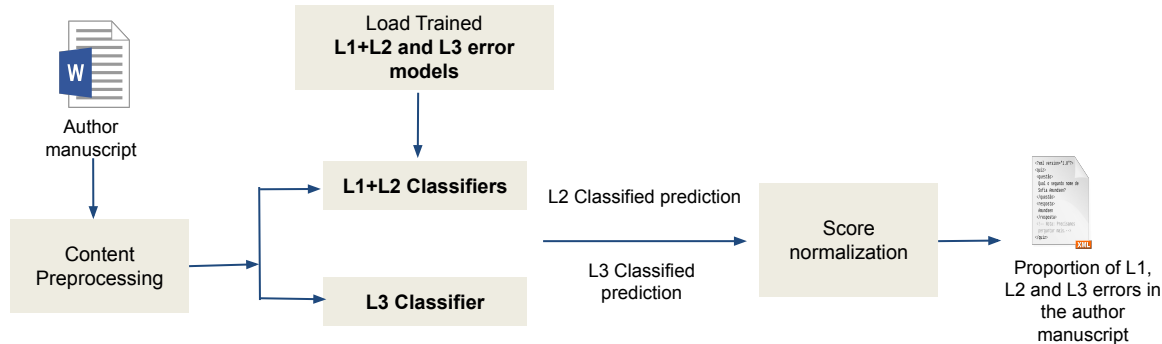

Academic book summarization result

Language Assessment Tool

Before even delegating author manuscript for editing to copy editors, identify the extent of spell check, grammar, structuring editing required for the document.

Quickly analyze the entire author manuscript for any level of error possibility and predict what type of editorial work is required on the same, whether spell check, grammar check or structuring/rewriting required.

3 standard error classification considered for our use case:

- L1 level - Spell or punctuation error
- L2 level - grammatical error
- L3 level - Content ordering, rewriting, and structuring required

Author manuscript

Content Preprocessing

Load Trained **L1+L2 and L3 error models**

L1+L2 Classifiers

L3 Classifier

L2 Classified prediction

L3 Classified prediction

Score normalization

Proportion of L1, L2 and L3 errors in the author manuscript

Multilingual transfer desk agent

Automated analysis of customer email communications, and predicting severity of issue and customer churn rate.
Help organisation to focus on critical and high priority issues, by analysing customer communication sentiment and pending issues from emails.

Approach

1. Parse customer email communication and communication chain.
2. From communication, identify customer sentiment & issue and highlight the potential severity of the issue.
3. Based on identified issue and severity, redirect the issue to the respective agent for quick resolution.
4. Move any pending or continuous issue chain for high priority.
5. Enable process to have more one time resolution strategy.
6. Focus more on similar issue identification and root cause analysis.
7. Prioritise customer and their issues, based on sentiment tone.

Email Classification Service

Classify client emails from one email box to two groups, research email and non-research email communication/application.

Approach

1. LSTM neural network based classifier built to first classify emails into research and non-research email type.
2. Then each research emails and attached document are further classified into sub-categories depending on the specific research type.
3. Attachments where compared based on document similarity and closeness to the sub research topic.
4. Emails are then moved into research/non-research boxes and then to respective sub-research category.

Result

ML accuracy : 88.7% and along with explicit rule addition, final accuracy 96%

Automated call and email analysis for Investigations

Audio recordings of the call agents with alleged misconduct were key to multiple cases, where it found emails and audio recordings that were evidence of rigging.

Approach

1. Technology based speech analytics solutions, leveraging insights from computational linguistics and machine learning along with strong regulatory expertise, can far outpace traditional, manual review of audio data by human listeners and offer unique new opportunities in regulatory and discovery applications.
2. Ensembelled **NLP + Audio** approach of combining below two:

   Phonetic search , the algorithm determines the phonetic closeness between a search term/defined pattern and a raw audio stream. On the basis of a predetermined confidence threshold, the system then outputs a hit report listing the timestamped occurrences, if any, of each search term/pattern item in each call.

   Speech recognition system converts a raw audio stream to a hypothesized natural language transcript, and any searches for specific terms are subsequently run over the resulting text transcript.
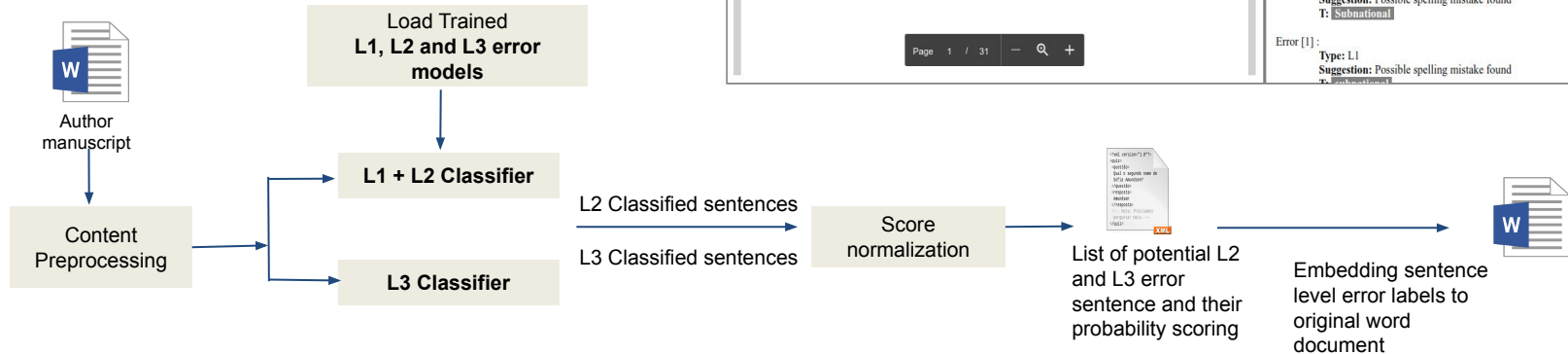
Outcome for a collection firm-

- 26% increase in agent compliance
- 66% reduction in quality monitoring costs
- 32% complaints reduction
- 86% collections revenue

Contextual Copy-Editing (Grammarly alternative)

Using **Artificial Intelligence** and **natural language processing** to process and score the language quality of a Journals and Books accepted into the copyediting workflow to determine a customised level of editorial intervention.

Based on the score we decide, if an article needs work - we should be copy editing it and if an article doesn't need work - we should leave it alone.

Author manuscript

Load Trained L1, L2 and L3 error models

Content Preprocessing

L1 + L2 Classifier

L3 Classifier

L2 Classified sentences

L3 Classified sentences

Score normalization

List of potential L2 and L3 error sentence and their probability scoring

Embedding sentence level error labels to original word document

Affiliation Structuring and Reference Structuring

References and Affiliations structuring using NLP.

Challenge Statement

1. Regular expressions based patterns repository.
2. 30,000+ reference patterns.
3. Difficult to maintain

Approach

1. Extract all reference and affiliation text in a word doc using ML AutoZoning and content extraction logic.
2. Pass extracted text references and affiliations to custom built ML model.
3. NLP probabilistic approach instead deterministic approach .
4. Used salient NER features for Machine Learning.
5. Additional Test corpus along with pattern recognition for entity extraction.

Structured affiliations

```
<author>
    <persName xmlns="http://www.tei-c.org/ns/1.0"><forename
type="first">Chengji</forename><surname>Shen</surname></persName>
    <affiliation key="aff0">
        <orgName type="laboratory">State Key Laboratory of Hydrology-
Water Resources and Hydraulic Engineering</orgName>
        <orgName type="institution">Hohai University</orgName>
        <address>
            <settlement>Nanjing</settlement>
            <country key="CN">China</country>
        </address>
    </affiliation>
</author>
```

Structured references

```
<listBibl>
    <bibl>
        <author>Gavazzi G., Piertini D., Boselli A., Tuffs R.</author>, <date>1996c</date>,
        <title level="j">A&amp;AS</title>, <biblScope unit="volume">120</biblScope>,
        <biblScope unit="page">489</biblScope>(<note>Paper I</note>)
    </bibl>
</listBibl>
```

Automated call auditing to enhance customer experience in a contact center

Call recordings can become a gold mine of rich insights about customer satisfaction, customer churn, competitive intelligence, service issues, agent performance and campaign effectiveness.

But the sheer volume of phone calls exceeds the contact center's ability to manually review and analyze them. Manual review can process only a fraction of calls using unsophisticated analysis.

Approach

1. Advanced audio analytics algorithm, a language independent solution predicts the individual agent's tendencies by linking speech patterns to personal characteristics like intonation, pace, emphasis focusing on prosodic speech parameters (non content based).

2. Calls are typically reviewed for agent quality, compliance, risk, customer satisfaction, and churn potential.

3. It searches for key phrases including cancel, unsubscribe, remove, stop my service, too expensive, cheaper option, and very unhappy, to track customer churn potential. When these words are mentioned, the call can be graded appropriately.

4. Conversational analytics bring out the tone or emotion of the conversation and highlights from the agent's responses. The solution can automatically discover and analyze words, phrases, categories and themes spoken during calls to reveal rising trends and areas of opportunity or concern.

5. In Depth analysis of agent customer interaction, to extract call specific critical parameters. Automatically detect and redact critical PCI (Payment Card Industry) details.

Outdoor adventure company-
- 82% increase in customer satisfaction (CSAT) score level.
- 100% increased targeted coaching
- 5% increase in close rate
- 18% increase in net promoter score
- 12% first call resolution (FCR)

<u>Technology Scouting</u> - recommend upcoming booming technologies in automotive industry

The client is a Technology Scouting company based in Tokyo (Japan) and California (USA). The company offers various manual scouting services to their different automotive clients need around the globe and across multiple departments. The ideal manual technology scouting cycle takes 3 months to 18 months for an automotive industry.

Challenge

1. Connectivity to the business - Multiple groups responsible for operating that radar
2. Identifying the right areas to scout - regional limitation & lack of skill
3. Time to scout - time gap b/w technology scouting and initial development

Approach

1. Digital sources included, research articles, blogs, company websites, scholar articles, online video series, holding briefing sessions reports, etc.
2. NLP and deep learning based multiple models were developed for below steps: Research topic extraction -> Research topic refinement -> Topic related meta info extraction -> Segmentation and relationship extraction ->Save data to data lake.
1. Return the structured responses upon search to executive & business leaders

Result

1. Passively waiting for technological developments is no longer required. Platform enables screening existing and emerging technologies in order to secure consumers competitiveness and innovative ability. The technology scouting platform can now serve as an early warning system for relevant technological changes.
2. Time to scouting reduced from ' 3 months - 18 months ', to 2 weeks - 4 weeks for multiple departments.

<u>Financial risk prediction</u> from external web sources + internal transaction data

For a new product, RISK related to its release are not always know from internal process flow data, but even from an existing product reviews, competitor new product release announcement, negative company feedbacks/news, similar product release in market, all such factors plays a vital role in computing the final risk related with a new product release..

Challenge Statement

1. Industry & product specific risk information are available all across the web (competitor website, public blogs, news platforms, product review websites, etc).
2. Manual extraction of ever changing financial or product related risk information is impossible

Approach

1. Use the industry relevant news and product review-source-platform urls for content crawling
2. Remove noisy and repetitive raw contents using core NLP based raw content scoring technique
3. Extract relevant entities (product SKU, brand, location, competitors reference), new topics, relevant product feedback/review statements and likes.
4. Structure individual url contents into a group, and pass it for risk assessment.
5. Record the potential risk factors, related custom score in data based for alter search.

Result

1. Potential product risk factors are now collected and populated to the company Chief-Risk Officer on daily basis.
2. Reduced the risk related information extraction from 8-10- days to less than 1 hour.

Banking loan fraud detection from customer call behaviors

Short 3 months POC (proof-of-concept)

1. Fraudsters use many approaches to trick call center agents and retail associates. Most don't even know they've been a victim of fraud until after the incident is over.
2. Built business rules and voice patterns to identify fraud behaviors, conversations, and even detect when a caller is faking or lying, so that steps can be taken to mitigate the risk.

Approach

1. Analyse archived fraudulent calls.
2. Identify textual (identify critical keywords) and voice patterns
3. Create potential fraudster scorecard, using - phrases, keywords, and voice and speech features.
4. Using business rules and thresholds, flag a potential fraud conversation.
5. Realtime assistance can be build in future, after analyzing a lot of such samples and patterns.
6. 56% overall accuracy (model + rule based) obtained in initial POC, after testing on limited audio samples dataset.

Automated zoning searchable PDF (section identification and tagging)

PDF content extraction starts with manually zoning all the sections (like title, images, sub-titles, page number, paragraphs, etc) in a PDF. Then it goes for section by section content extraction

- Using document text structure meta information and neural network based image processing, we have automated the entire manual zoning process.
- To verify the searchable PDF zoing results, we have used scanned/image PDF results.
- At the end we have generalized the PDF to xml conversion & auto-zoning, by treated 'the text problem as an image problem'.
- Approach is to identify different text sections and their section types (title, subtitles, tables, borderless tables, images/graphs, equations, abstract, references, bullet points, and more), along with bounding box and related text (from searchable PDF. For image input, use OCR to convert image bounding to text)
- Generating final structured zoned and tagged XML of searchable PDF and scanned/image PDF.

Auto sectioned sample PDF output

Auto Proofreading

Using machine Learning Convolutional Neural Network, automate the document proofreading by highlighting and validating the document layout & alignment errors.

Approach

1. Samples collected from live and created Annotating the error labels into PDF pages.
2. Split and convert PDF to multiple images using GhostScript.
3. Preparing the XML format of annotated labels and box values from PDF using.
4. Use config file, training corpus image dataset, text file (converted from XML) and set the epochs & epoch length depends upon training corpus size.
5. Get the Trained Model, configuration pickle and training weight file after training.
6. Save Model for real time inference.

Invoice content extraction

Using Machine Learning OCR, NLP based named entity extraction, layout extraction and rule engine to extract content from PDF Invoice into XML.

Approach

1. Content preprocessing.
2. Contour based text region identification.
3. Extracting text content using OCR and passing it to NLP engine for entity extraction.
4. Trained NLP model extracts content from the input text and returns key value pair.

5. Extracted Content is feed into predefined layout.
6. If no content extracted or doesn't meet any of the saved layout, then allow operator to manually label the content and save the new layout to the layout repository.
7. Extract Invoice content into XML output.

Smart OCR (computer vision and NLP)

Template based form extractor OCR. Extract handwritten text from bank form scanned image (any form scanned copy), using template matching, individual box extraction and OCR. Train your own character and alphabet OCR with pytesseract.

Approach

1. English handwritten character by charecter training dataset generated.
2. Neural network model trained on handwritten English characters.
3. If the number of forms are fixed, then create multiple templates with the position of the content with tags. Else use next step with image processing.
4. Document preprocessing using image processing logics - binarization, hugeline detection and elimination and denoising.
5. Using NLP to auto-complete the missing details or correct the spellings, identity names and address for structuring.

Alt-Text generation

Neural Network based, generative model built for captioning images in natural language, English. Keras with TensorFlow is the application setup for this project. Sample size of 20k labeled images were used as a training set for the model development.

The overall process consist of three core components :
1. **CNN encoder model** (A pre-trained CNN is used to encode an image to its features and also pre-encoded each image to its feature set for high performance and speed).

2. A **word embedding model** (First tried with pre-trained word embedding model over word2vec models and also explicitly trained and embedding model that takes a word and outputs an embedding vector of dimension (1, 128)).

3. **CNN decoder model** (It takes the image vector and partial captions at the current timestep and input and generated the next most probable word as output with the LSTM).

Caption text : white crane flying over the water

Image tags (prediction confidence %):
water — 79.12%
crane — 73.76%
over — 57.50%
flying — 47.04%
the — 36.27%
white — 23.83%

Content Preprocessing → **CNN Image Encoder** (Load Trained **Encoder Model + Weight**) → Encoded Image Vector passed in sequence at each timestamp → Sequential RNN on LSTM network (Load Trained **Word Embedding vectors**) → Generate next possible sequential word output at each time stamp. → **Beam Search** → Return top scored caption

Ad Entity Extraction (NLP + OCR + web sources)

Extract text and image properties from a given advertisement image. Use case was for a large advertising agency, where the task was to extract Image properties from the Image.

Approach

1. Built custom image processing ML model to identify text and non-text entities from a given image.
2. Using Natural Language processing, in combination with business rules to get the most out of the extracted text, like event information, location, time, point to note and disclaimers etc.
3. Band logo and related detail extraction, like name, model/type, and person as an identity, etc.
4. Using external web sources to cross validate the extracted information.

{"brand": "Denso",
"text": {"t1": "Who Takes Climate Control",
"t2": "DENSOA/CCompressors:Chillingly Precise.The compressoris the heart of the A/C system, and over 80% of the A/C compressors[DENSO makes are DE onto production cars. Automakersknowtheydonthavetosweat becausetheycantrustDensotomakehigh-outputcondensersthatprovidequietande fficientperformanceeextendeddurability,a ndunmatchedreliability.
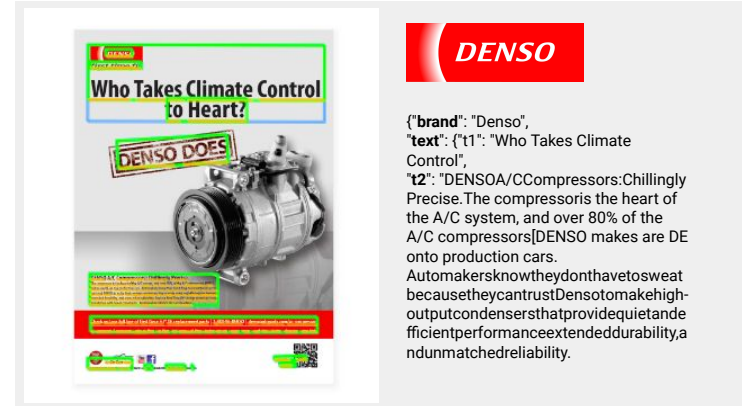
Healthcare customer leakage risk prediction from agent-customer audio recordings

Healthcare providers face a lot scrutiny every day. Speech analytics lets providers uncover hidden reasons for setbacks, make improvements, meet regulations, and achieve a better patient experience. It means predicting customer future complaints caused by misunderstandings, process issues or unrealistic expectations.
Healthcare are leveraging audio analytics to generate insights into customer needs.

Approach

1. Voice and speech analytics in healthcare, can automatically transcribe and score every patient interaction to identify relative compliance risk and give next-best action to quickly address issues during the interaction.
2. Advanced audio analytics, speech analytics language model, sentiment analytics and phonetic search automatically piece together a full conversation and identify common, trending, and hot topics.
3. The solution works by building keyword and key phrase search definitions within a speech analytics solution, and as the calls are processed, they are categorized by the keywords and phrases that define a search, which is customized for every customer. This information can be used to help improve clinical performance and marketing effectiveness as well as provide better customer service.
4. It helps predicting customer future complaints caused by misunderstandings by searching categorized keywords/phrases, and also process issues or unrealistic expectations.
5. When patient representatives know exactly what areas they need to improve, they are usually able to make adjustments on their own and deliver a better experience to the patients.

US healthcare company-
● 42% achieved regulatory compliance through speech analytics
● Processed 100% of the customer - agent interaction for full quarter.
● Manual quality check along with the automated solution, helped to automate the entire process by 70+%.

Generate cross sell opportunities for a leading UAE based insurance firm

A large Insurance provider experiences a large volume of calls daily in its call center.

The client was looking for opportunities where the caller is a potential upsell or cross-sell candidate. The idea is to pitch the right product at the right time to the right customer and checking whether the agent is approaching that appropriately in terms of their dialogue, their persuasive selling ability and how they overcome objections. Large volume of calls make data analysis through manual sampling and listening virtually impossible.

Approach

1. Speech analytics language model and machine learning can be positioned as a revenue generation tool. Industries can achieve their revenue goals on the basis of speech analytics and the value it provides in product cross-sell and up-sell sales opportunities. Use it to extract insurance (can also work in banking) critical fields from a conversation transcript. Critical fields like, product name, model number, call type, location, area code, design type, or any other important entities specific to a business.
2. Voice analytics can be used in addition to speech analytics to focus on the content of customer conversations (how it is said) to derive the context, tonality of conversation pertaining to product or services.
3. Sentiment analytics focuses on the current disposition of the customer to derive the happiness factor of the customer (positive, neutral or negative). It considers any interaction, demographic, or engagement life cycle information about the issues that matters most to the business.
4. Customer Segmentation – Customers grouping based on extracted demographics data, buying behavior or some other patterns which are further used in marketing.
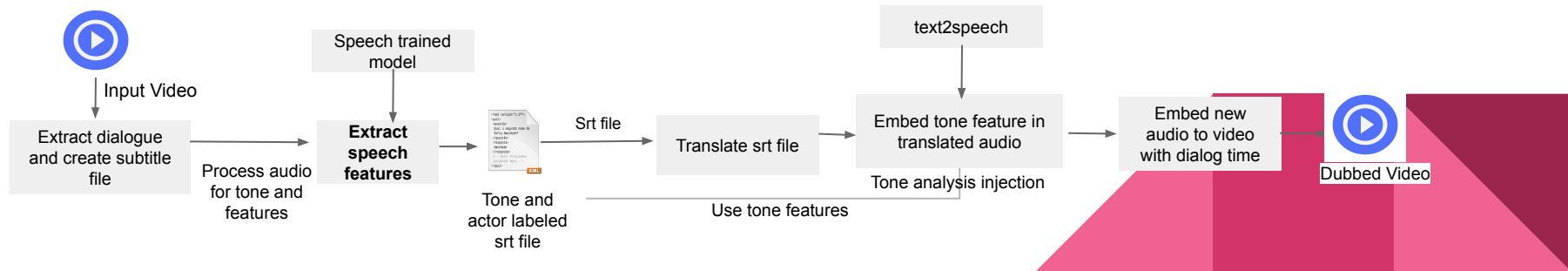
UAE Car insurance firm-
- 26% increase in cross sell
- 100% increased targeted coaching
- Complex customer demographics and spoken business critical fields get extracted from each conversation for call tagging and categorization.

Auto-Dubbing 14000+ hours of tv series in multiple Asian languages

Automate solution for dubbing 14000+ hours US english archived tv series recordings into multiple Asian language (Vietnamese, Hindi, English, Russian, Indonesian, Tamil, Telugu and Bengali)

Approach

1. Video auto subtitleing.
2. Video emotion extraction (using emotionML).
3. Audio feature extraction.
4. Tone, pitch, intensity, analysis.
5. Subtitling translation to another language using 3rd party transcribe service.
6. Use text2speech library to convert translated text to another language speech.

6. Embed original tone intensity features from original audio to new audio, also embed video emotion features to make the final audio output look more realistic.
7. Embed new audio to original video file to generate dubbed video file.
8. Use minor manual modulation (if needed) to sync the final dialogue speed and timings

Input Video

Extract dialogue and create subtitle file

Process audio for tone and features

Speech trained model

Extract speech features

Tone and actor labeled srt file

Srt file

Translate srt file

Use tone features

text2speech

Embed tone feature in translated audio

Tone analysis injection

Embed new audio to video with dialog time

Dubbed Video

Cervical cancer detection (WIP)

World's first multi lingual cervical cancer detection application assisted with an external zooming device and AI image processing for auto detection of cancerous regions in the image

Luein analytics
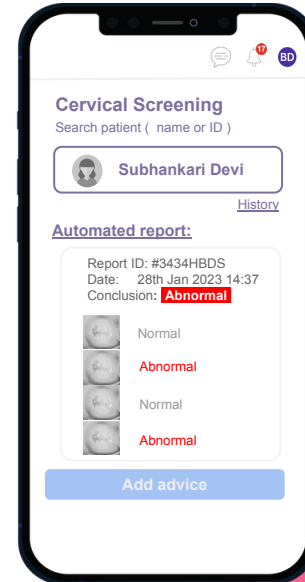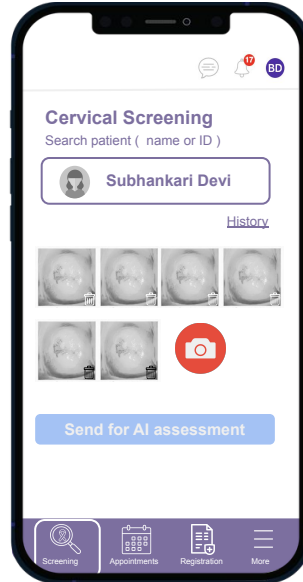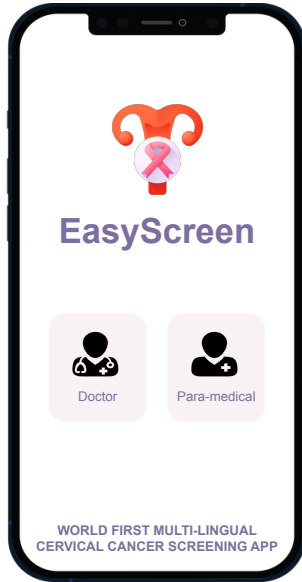
Solving AI Challenges Swiftly and smartly

Harendra Singh
AI Architect & Strategist
Bangalore, India

Dr. Priyo Chatterjee
AI Strategy Consultant
New York, U.S.

Vibhor
AI Strategy Consultant
Bangalore, India

Dr. Shiv Shankar
Chief Strategic Advisor
Delhi, India

Dr. Kirran Kandam
AI Psychology Practices
Ph.D. in Clinical Psychology

Dr. Navneet Agarwal
Chief Advisor
Gwalior, India

Dr. Bhawna
AI Strategy Consultant
Ph.D. in ML

Emml
Senior Data Scientist
Greater Leicester Area, UK

Vidyaratan
Senior Data Scientist
Bangalore, India

Sulabh
AI Strategy Consultant
Delhi, India

Vidit Agarwal
Senior Data Scientist
Bangalore, India

Awadhesh
AI Architect
Bangalore, India

& 30 more….

We are a community of friendly, open-minded, collaborative and professionally growing people. Our team is made up of some of the brightest minds in statistics, math, data science, and, technology and we have obsessed with how AI is defining the customer experience.

Recognitions

11
Client Patent Assistance

18
Publications

20+
Workshops

What clients says about us

"Working with Luein Analytics on AI use cases was (and continues to be) an amazing experience. We worked closely with Luein Analytics to execute on our shared vision of making customers successful through the adoption of AI that accelerates enhanced decision making around an ever evolving risk and compliance landscape. They are extremely helpful and professional, they made everything so seamless for us, and we look forward to continuing the journey."

- **Naveen Kumar TV (Managing Director, 4CRisk.ai)**

"Harendra Singh is a gifted enthusiastic social entrepreneur with a rare combination of patience, spunk, and research skills. Harendra is a AI strategy advisor to us for many products development in Artificial Intelligence (AI) and Data Science for APAC regions. I happily recommend him here on LinkedIn, and do so in person to my clients whenever I have the opportunity"

- **Srinivasan (Director, kamlaxglobal.com)**

Thank You

Contact us:

LUEIN ANALYTICS RESEARCH AND CONSULTING (OPC) PRIVATE LIMITED

INDIA: #677, 1st Floor, 27th Main, 13th Cross, HSR Layout, Sector-1, Bangalore, Karnataka, 560102

MALAYSIA: CoPlace2,2260 Jalan Usahawan 1, 63000 Cyberjaya, Selangor Darul Ehsan, Malaysia.

Contact: +44 20 3289 9877 (**UK**), +91-8792180457 (**IND**) and +1 (650) 3537693 (**USA**)

Email: contact@lueinanalytics.com & harendra@lueinanalytics.com