



AIA SHI

ARTIFICIAL INTELLIGENCE
AS
HUMAN INTELLIGENCE

General Problem

- Global Context
 - Uncertain
 - Rapidly evolving
- Companies need to:
 - Extract added value from data
 - Anticipate risks and opportunities —> Detect weak signals
 - Construct scenarios
 - Make strategic decisions

Solution

Solution

- Unsupervised Machine Learning
- Natural Language Processing
- Modular workflow

Features

- Uncover weak signals
- Discover and track emerging trends
- Support scenario building
- Profiling

Benefits

- Increase efficiency of agents
- Understand arising risks and opportunities
- Anticipate market disruptions
- Improves environment awareness



Case: HR Analytics

- How will tech market evolve in the next 10 years
 - Anticipate market changes & prevent employee churn (risks)
 - Prepare for employee training & recruitment (opportunities)
 - Align new technologies with corporate strategy (decision making)
- Weak Signals
 - Different from mainstream trends
 - Refers to event that could become a trend in the future
 - Tech. market changes



Requirements

- Interpretability of machine learning models
 - No Deep Learning
- Lean model
 - Easily run on in-house infrastructure
- Interpretability of results
- No threats to workforce
 - Automation, substitution of HR officers not intended



Detecting Weak Signals

- Data source
 - 100K job ads (Indeed.com)
 - Unstructured texts
 - Diverse language
 - EN, FR
- Privacy?

IT Newbies 2023 - Data Scientist



Euroclear ★★★★★ 28 avis

Belgique

Créez un compte Indeed avant de continuer sur le site web de l'entreprise.

Continuer pour postuler



Division: Group Digital Capabilities (GDC)

Within the GDC Division, the AI center of excellence supports the needs on data science/machine learning from all the entities of the Euroclear Group. As a competency center for analytics, the team helps to transform data into insight using techniques such as natural language processing, predictive modeling, mathematical optimization or network analytics.

To help you seize all the opportunities available when launching your career at Euroclear, we established the IT Newbies Graduate Program designed to help unique young talents join our dynamic international teams with extra-guidance and support at all levels.

Your role

We are looking for junior profiles for the role of data scientist starting their learning curve within our team.

As **Data scientist**, you will create innovative solutions that rely on data science/machine learning and experiment new technologies to bring value to Euroclear as a data and insights-driven company.

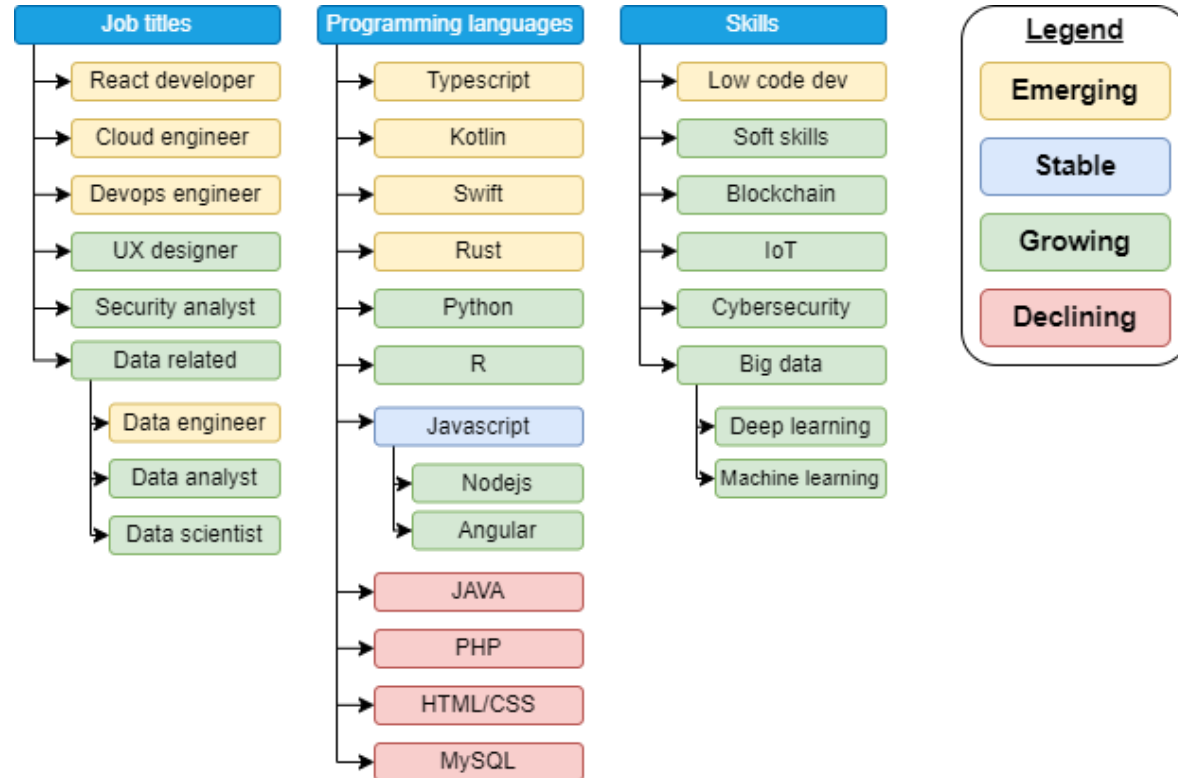
In order to reach these objectives, you will:



Approach

- Unsupervised learning
 - Hierarchical & Temporal Topic Detection
 - See [Poumay & Ittoo \(2023\) for algorithm](#)
 - Applied to jobs ads
- Output
 - Weak signals (emerging trends) vs. mainstream trends vs. declining topics
 - Hierarchy
 - Relationship between general (parent) & more specialized (children) topics

Result :Landscape of the future tech market



Hierarchies & temporality provided fine grained analysis for strategic decisions



Use-Case: OSINT

- Open-Source Intelligence
 - Gathering, analysis open-source information
- Application
 - National Security/Defence



OSINT

- Proliferation of social media/messaging platforms
- Security analysts teams
 - Overwhelmed by volume of information exchanged online
 - Distinction actionable intelligence vs. noise
- OSINT automation as solution



Requirements

- Interpretability of machine learning models
 - No Deep Learning
- Lean model
 - Easily run on in-house infrastructure
- Interpretability of results
- Tweets as datasource
 - Reflection of citizens' conversation



Risks & Opportunities

- Risks

- Privacy infringement
 - But warranted here for security?
- Limited resources of analyst teams
 - Gradual replacement by AI applications
- AI applications lack cultural, social knowledge
 - Difficult to define notion of “threat”

- Opportunities

- Ability to scour wide-ranging online sources
- Real-time analysis
 - Threat detected and mitigated instantaneously
 - Avoids potentially drastic consequences
- Human analysts can focus on more value-added activities
- Significant cost-savings

{"created_at":"Mon Oct 30 12:48:36 +0000 2017","id":924981362225954816,"id_str":"924981362225954816","text":"\u92fc\u9244\u306e\u9b54\u5973\u30a2\u30f3\u30cd\u30ed\u30

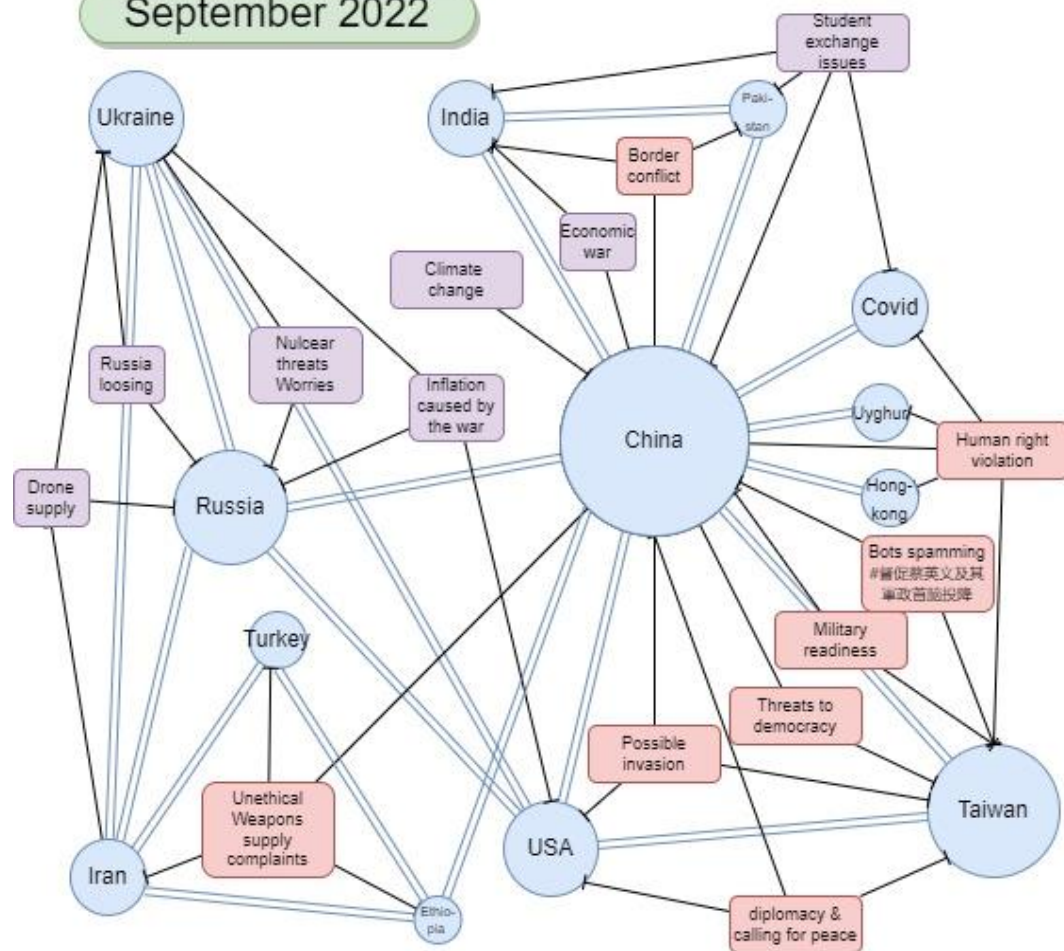
{"created_at":"Mon Oct 30 12:48:36 +0000 2017","id":924981362741870592,"id_str":"924981362741870592","text":"RT @flipsidefeels: You can Download your copy here [bugo](#) -

{"created_at":"Mon Oct 30 12:48:36 +0000 2017","id":924981363287232512,"id_str":"924981363287232512","text":"RT @Vrairapfrançais: Orelsan a vendu 96 974 [exemplaires](#) de

{"created_at":"Mon Oct 30 12:48:36 +0000 2017","id":924981363266281472,"id_str":"924981363266281472","text":"RT @AmbroiseDebret: Aaron [Agius](#) on Digital [Nomadism](#), worki

{"created_at":"Mon Oct 30 12:48:37 +0000 2017","id":924981363828146176,"id_str":"924981363828146176","text":"RT @SadoHakase: \u6deb\u8fb1\u306e\u5973\u4f53\u30aa\u30fc

September 2022





Use-Case: Insurance Churn

- Longstanding issue in marketing
- Case of Insurance company
- Weak signals
 - Determine who is likely to churn
 - Earliest possible in customer journey
 - NBA (next-best-action)
- Not an NLP use-case
 - But relevant to many organizations



Churn

- Complex tasks
 - Several reasons to churn
 - Different values of customers
- Requires human expertise
 - Actuaries (risk modelling)
 - Marketing officers
- Can still be automated?



Risks & Opportunities

- Risks

- Lack social/cultural knowledge & likelier to discriminate
 - Higher premiums based on address, education level
- Automation threats to "white collar job" (actuaries, marking officers)

- Opportunities

- Cost savings
 - More expensive to acquire new customer
- No loss to recurring revenues (subscriptions)



Data

- Customer records
 - Size: approx. 500K customer records
 - Described by approx. 50 features



Data

- Customer
 - History
 - House hold view
 - Civil status
- Product
 - Type
 - Vehicle type insured
- Insurance claim
 - Number of claim
 - Household view

...



Real-life Data Challenges

- Churned customers still appearing as "existing customers"
- Class imbalance
 - Much more non-churners than churners
 - 10: 1 ratio
 - Biased prediction
- Missing values



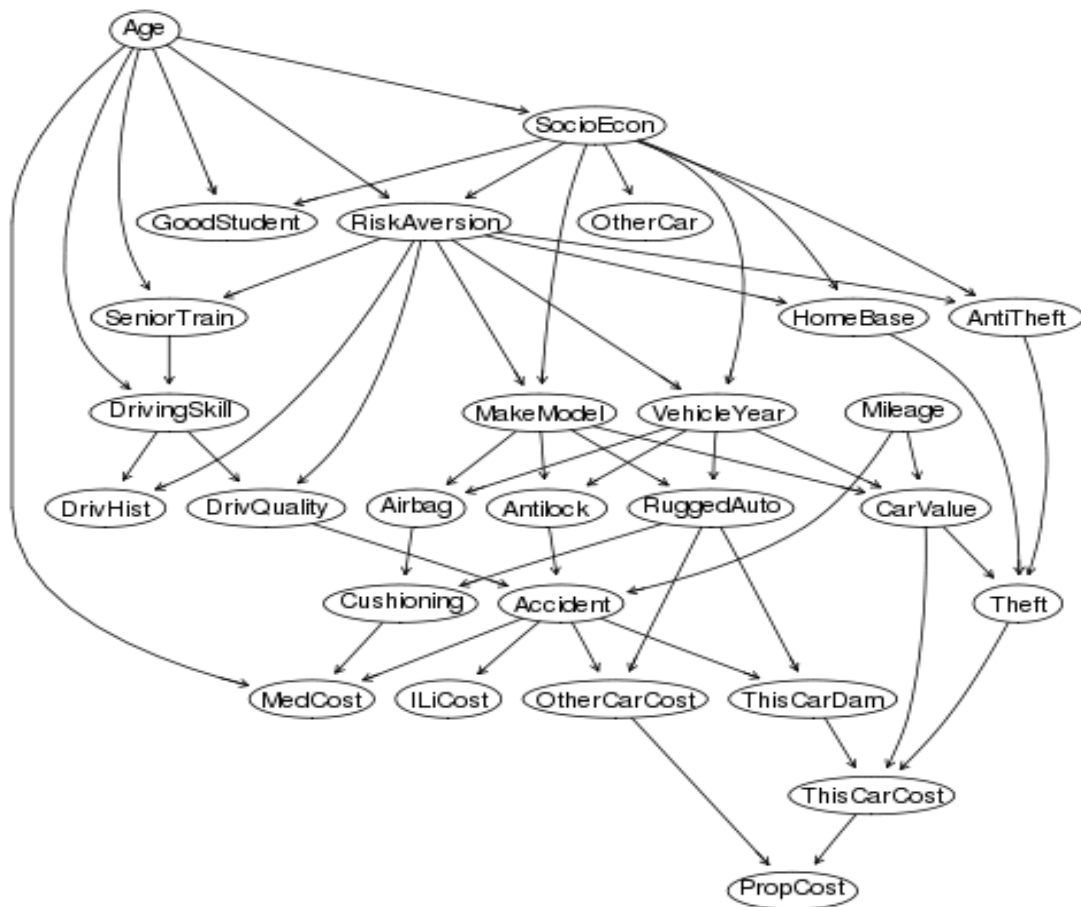
Requirements

- Interpretability of machine learning models
 - No Deep Learning
- Lean model
 - Easily run on in-house infrastructure
- Interpretability of results
- Strong *focus on removing "bias"*
 - Data imbalance responsible for biased prediction
- Strong *focus on trustworthiness* of predictions
 - Estimate uncertainty in predictions



Approach

- Bayesian belief networks
 - Easily interpretable
 - Ability to understand variable's influence
- Credal sets, ensemble
 - Estimation of uncertainty






Key Results

- Estimate for each customer:
 - VIP customer
 - Customer lifetime value (adjusted over customer journey)
- Estimate risk associated with customers
 - Low risk vs. High risk
- Low risk
 - Next best action if churn
- High risk
 - Liabilities, no reason to keep

Use-Case: Customer Service Automation

- 
- Heath insurances ("Mutuelle")
 - Customer service agents
 - Increasing number of requests for informations, questions
 - Diverse topics
 - Reimbursements
 - "BIM"
 - Eligibility criteria



Customer Service Automation

- Large number of disparate sources
 - Customer emails
 - Call-centre transcripts
 - Web-pages
 - Internal documents/memos
 - Legal documents & contracts
- Information integration



Customer Service Automation

- Challenging endeavor
 - Increase in number of data sources/sizes
 - Increase in customer queries
 - Complex questions
- Overwhelmed agents
- Dissatisfied customers
 - Lack of clear information
 - Uncertain about eligibility criteria
 - Late reimbursement



Business Objective

- Automatic question-answering system
- Able to engage in dialogues/conversation
- Guidance through customer journey stages
 - Registration
 - Claims submission
 - Reimbursement
 - Document verifications
 - Advice on contractual, matters



Solution

- Conversational agent
- ChatGPT?



ChatGPT as Solution

- Risks

- Issue of confidentiality
 - Even more with medical, health-related data
 - (Recall Privacy slides earlier)
- Automation threat to customer-service agents
- Hallucination
 - Wrong guidance provided to customers
 - Legal implications

- Opportunities

- Analysis of pertinent information from wide-variety of information sources
 - Precise, relevant answers
- Customer-agents can focus on more complicated cases
 - Providing personalized, in-person interviews
- Overall: increased customer & employee satisfaction



Requirements

- Ability to integrate&analyze different data sources
 - Emails, call transcripts, internal documents, legal contracts
- Anonymisation of data
- Reliable, trustworthy answers
- Interpretability less of a concern



LLMs

- LLMs as promising solution
- GPT-x (e.g. GPT3.5 or ChatGPT)
 - Aligns with the aim and requirements
 - Relatively mature & well-known
 - Works quite well (besides hallucination)
- But no GPT-x...why?
 - In-house deployment difficult (impossible)
 - Only via OpenAI API calls via code
 - Lack of understanding of what happens to ingested data
 - Highly confidential medical data
 - Proprietary
 - Relatively heavy model
 - Training, fine-tuning computationally expensive

Open-Source LLMs

- Two models of interest
 - Falcon-70B
 - Llama-65B

	Llama	Falcon
Datasets	Refinedweb, Redpajama, GLUE, OSCAR, Wikipedia, Platypus, Puffin, Tiny series, Common Crawl, BookCorpus, CC100, Reddit Gigaword, SuperGLUE, OpenWebText2	Refinedweb, Pile, PaLM OSCAR, C4, CC100
Languages	English, French, Spanish, German, Dutch, Russian, Italian, Ukrainian, Swedish Portuguese, Romanian, Slovenian, Serbian	English, German, Spanish, Dutch French, Italian, Portuguese, Polish, Romanian, Czech, Swedish
Tokens	2 trillion (for all the Llama-2 models) 1 trillion (Llama-1-7B,13B) 1.4 trillion (Llama-1-33B,65B)	1 trillion (Falcon-40B), 1.5 trillion (Falcon-7B), 3.5 trillion (Falcon-180B)



Current Status

- Anonymization of data
- Fine-tuning over specific data
 - Emails
 - Call-center transcripts
 - Internal memos
 - Legal documents, contracts
- Defining evaluation benchmarks
 - Precision of answers
 - Truthfulness, reliability



General Conclusions Lessons Learn

- Most corporate applications favor simple methods
- Hard to accept complex methods
 - Deep Learning
 - LLMs
- Main desiderata
 - Interpretability of model & results
 - Privacy-preserving methods
 - Size, run-time
 - Integration with existing IT infrastructure
 - Truthfulness, reliability of results

Curious



- Reach out
 - [LinkedIn](#)
 - [Mail](#)
- Join our ExEd programme
 - [Digital Transformation, Innovation & Change](#)
 - [Créer de la Valeur Ajoutée avec le Machine Learning](#)