GAZPROM GERMANIA

The News Analyzer Project

Reading the News Programmatically

Agenda

- Introduction
- Intro to problem to be solved
- Why are we solving this?
- Solving Data
- Solving Processing
- Solving Analysis
- Solving Notifications
- Summary
- Tech Stack
- Q&A

Who am I/What do I do?

Stewart Ridgway – Development Team Lead in Data & Analytics

The Team

- Based in multiple offices (UK, RU)
- Mixture of Python and .NET developers supported with Quality Assurance testers, Business Analysts, DevOps, Database Administrators

What does the Team do?

- Provide assistance, support and software development for the Front Office trading desks e.g. BigData, Trading Platforms inclusive of: pre/post execution
- Design and Develop: Analytical modelling tools for basic through to complex modelling to assist in trading decisions

Data Overload

- Gazprom Marketing and Trading Energy Traders need data to make decisions around buying/selling of Natural Gas on the commodities markets.
- Every trader is exposed to a **significant volume of data on a daily basis.**

Problem: Makes it challenging to read all of the data and try to make a trading decision within a short space of time (Typically seconds)

To make matters worse:

- Energy Markets can be very sensitive to any ‘Event’ driven News
- Events can be: Geopolitical, Natural/Accidental Disasters, Government/Legal changes, Weather, Climate et al.

Reading the News – The Challenge

The Challenge

- To read the news around the world from millions of sources of which we need to identify important Trading related news that alerts Traders of an event.
- Provide a **Trade Signal** to the Energy Trader when something important is worth reading.
- Each news item **must be:** Read, Cleaned, Translated, Processed **<u>less than 1 second</u>**

Big Data

Big Data Processing daily statistics:

- Twitter – 550million tweets a day
- BBC News – 10k articles a day
- BloombergReuters/EIN Energy – 100ks notifications a day
- RSS feed – 100ks notifications a day
- Bespoke Sources – 10Ks articles a day

Challenges:

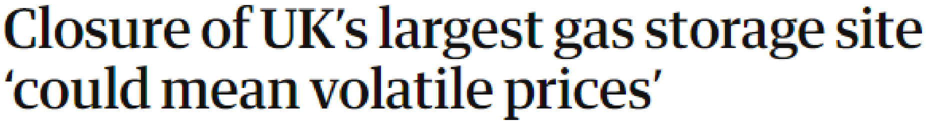
1. Different data types/formats
2. Frequency of data
3. Varying sizes of data

Why is the News important? – An Example

Rough Storage – Explain (Stores Gas under the Ocean because the UK has limited Natural Gas storage)

- In 2017 Rough Storage went offline due to cracks and failures
- Rough storage held the largest amount of Natural Gas in the UK.
- Most of the large Energy companies were exposed to holding Gas there and lost Gas
- The Gas market price became volatile.
- Reports and news were coming in slowly about continual updates.

<u>Some Energy companies were more aware of the issues earlier than other companies</u>



Closure of UK's largest gas storage site 'could mean volatile prices'

"5,000,000,000 cubic metres of Gas lost in one day"

What did Gazprom do to solve the problem?

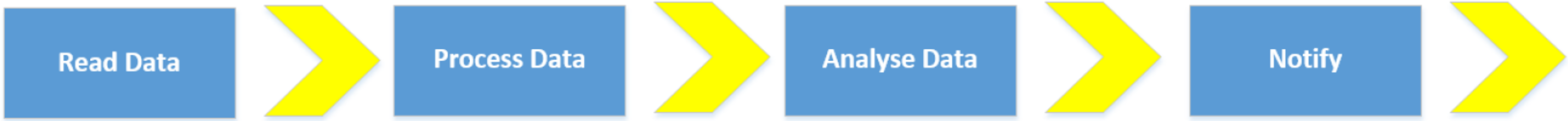
Gazprom Marketing and Trading were receiving news alerts from Reuters and Bloomberg

The Problem: Commodities Market, seemed to know before Reuters and Bloomberg about News incidents. How? Why?

The Approach:

- The project focused on sourcing data from multiple source including non-traditional news outlets
- Development of an application that could process large volumes of data and identify News-worthy items
- It needs to process each item in less than 1 second

Breaking down the problem



Read Data

Process Data

Analyse Data

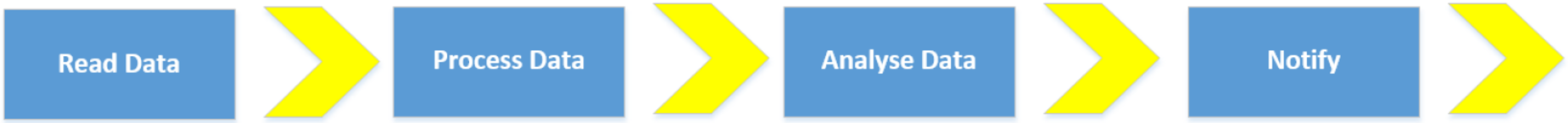
Notify

Reading Data - Problem

Given the challenges of consuming data – how do we consume it at high frequency?

The Problem: Read and process multiple format, multiple language data extremely fast

The Approach: Microservice techniques to scale with data



Read Data - Tasks

Tasks

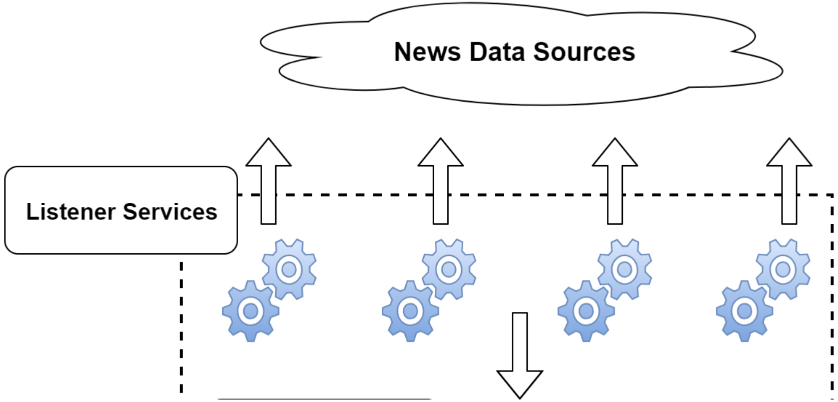
1. Listening to data changes
2. How do we technically handle large data pulses?
3. When does data change? (Frequency)
4. Handling multiple formats:

- JSON
- RSS
- YML
- HTML
- Text

Read Data - Solution

Solution: Break down News source into microservices:

- Each source has its own **Listener microservice**
- Each Listener knows what type of data it will handle
- Monitoring and frequency handled by spinning up a new microservice on demand

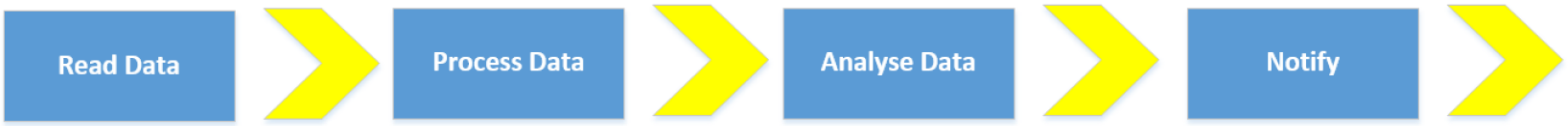


Process Data - Problem

Processing data is needed at a fast speed

The Problem: How do we translate, treat, clean, identify and categorise data at high speed?

The Approach: Microservice techniques to scale with data



Process Data - Tasks

Tasks

1. Check if data already processed (remove duplicates/re-tweets)
2. Translate to common language (English)
3. Use a dictionary to fix words (colloquial challenges)
4. Format data into a templated data set
5. Cleaning data strategies
6. Basic/Initial Machine Learning Analysis

Process Data

Yandex / .NET Libraries

- Yandex language Translation (translating 50+ languages instantly)

- Tagging/Entity detection and lemmatisation
  StanfordNLP.Core.NLP
  Link: https://sergey-tihon.github.io/Stanford.NLP.NET/

- Microsoft Cognitive Services (Text Analytics): “Key Phrases”
  Link: https://azure.microsoft.com/en-gb/services/cognitive-services/text-analytics/

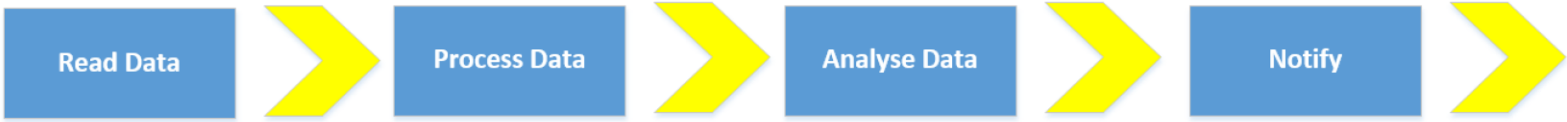
- TweetSharp (Read Twitter easily – recommend nuget package)
  Link: https://www.nuget.org/packages/TweetSharp/

Analyse Data - Problem

Making sense of the data we have cleaned

The Problem: How do we confirm whether the data is important and what is not?

The Approach: NLP techniques, Machine Learning, Supervised Learning, Categorisations

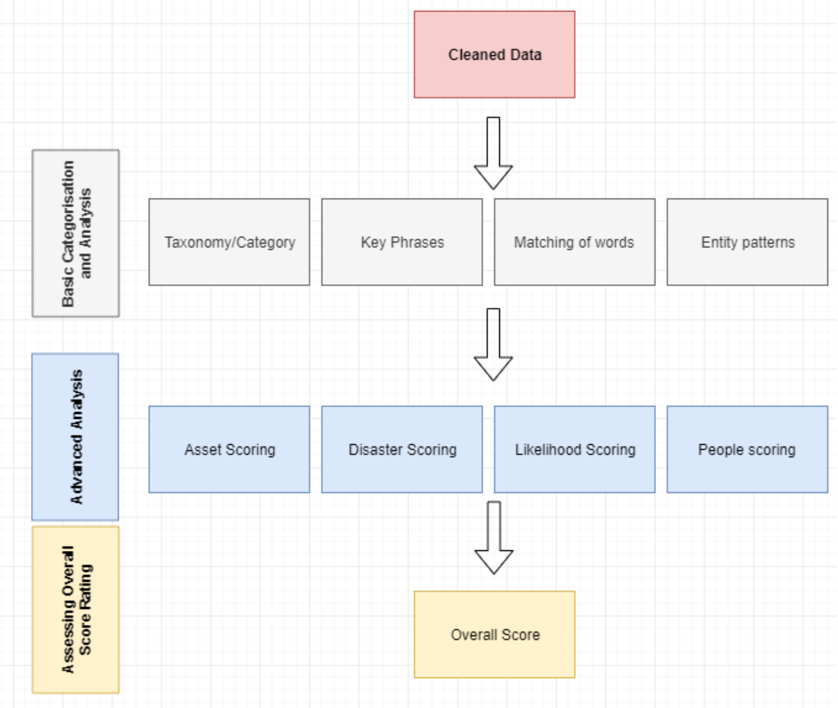


Analysing Data - Building Intelligence

Some of the challenges we faced:

- Processing News is good but how to measure success?
- Are there patterns in multiple news items that can confirm ‘Truth’
- Fake News?
- Traders may have different perspective compared to citizens
- Back-testing previous calculations/news
- What if the same News item appears again?
- How trustworthy is the Newsource?

Steps of Analysing data



Cleaned Data

Basic Categorisation and Analysis

Taxonomy/Category

Key Phrases

Matching of words

Entity patterns

Advanced Analysis

Asset Scoring

Disaster Scoring

Likelihood Scoring

People scoring

Assessing Overall Score Rating

Overall Score

Data has been cleaned/processed

Basic Analysis/Data Shaping

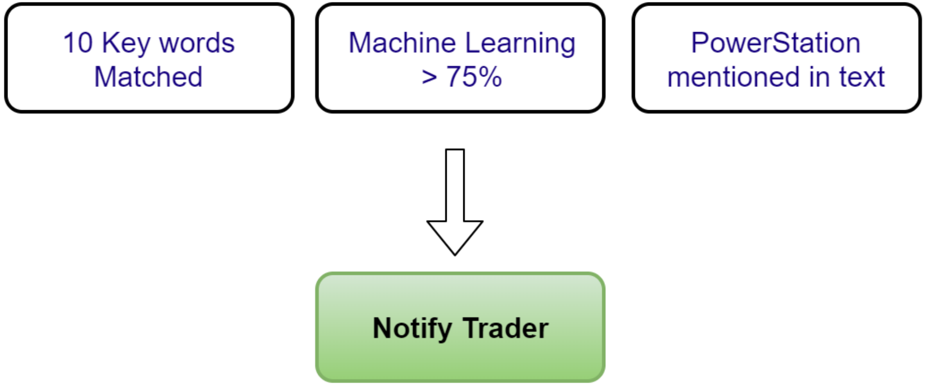
Adv. Analysis evaluating matches

Evaluating the value of the News Item

Deciding when to Notify a Trader

Analysing Data

It is a combination of factors that make a decision

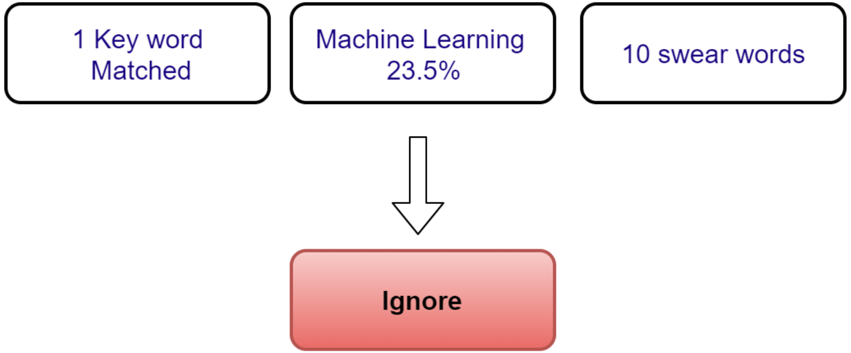


10 Key words Matched

Machine Learning > 75%

PowerStation mentioned in text

Notify Trader



1 Key word Matched

Machine Learning 23.5%

10 swear words

Ignore

Notifying the Traders - Problem

Notify

How to communicate all of this Analysis back to the Traders

The Problem: How to we let Traders know something happened?

The Approach: Email alerting, applications, data needs to be simple and easy to understand

Read Data

Process Data

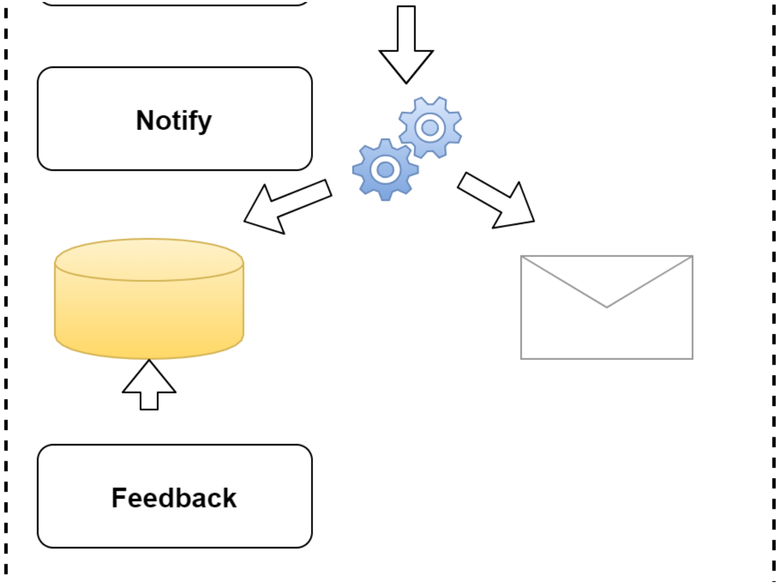
Analyse Data

Notify

Notify - Solution

Solution

1. Retrieve results of Analysis
2. Prepare data
3. Create email template to store data
4. Send notification to Trader
5. Consume Trader feedback

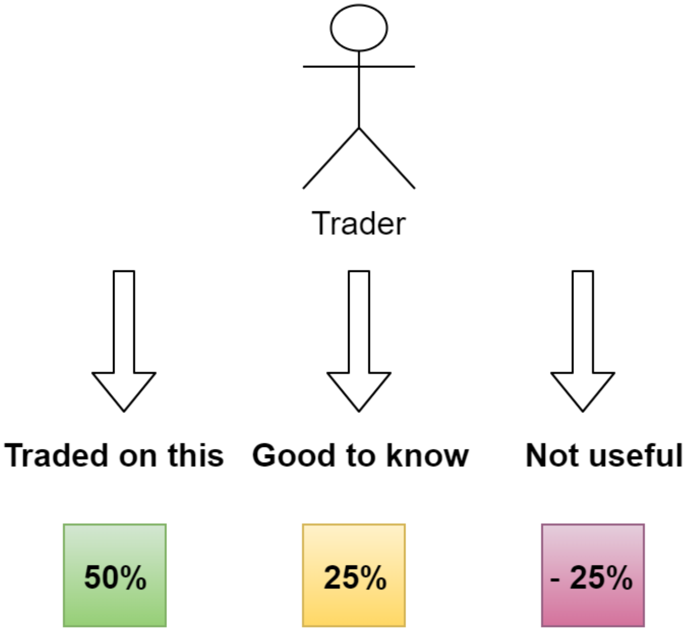


Trader Feedback

Notify

Not all News is perfect – How do you feedback?

- It must be fully automated!
- A simple system that people can give opinion on importance/relevance
- System takes the response and adjusts weighting of all words and data
- New News Alerts will use the adjustments



Trader

Traded on this	Good to know	Not useful
50%	25%	- 25%

.NET Libraries

- Standard .NET Core framework

- Experimentation now with ML.NET
  Link: https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet

Summary

- Reading the News programmatically can be a challenging concept but it is possible

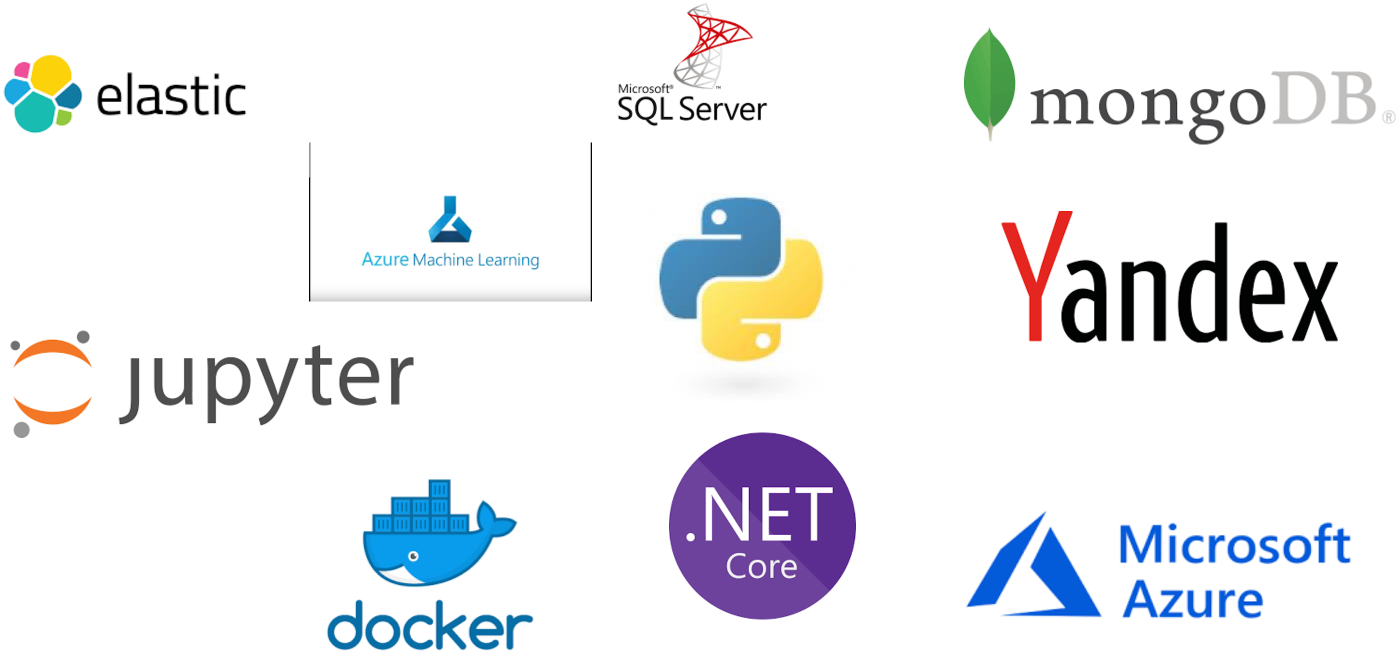
- .NET and MS Azure have played a large role in the project, Cloud technology helped to enhance speed of delivery and performance of the system

- Using many techniques in NLP a solution to identify and Read the News became an enjoyable challenge and experience
  - We also established what doesn’t work!

- Machine Learning is a sub-set of AI but is there really true-AI or is this too ambitious?

GAZPROM GERMANIA

News Analyzer - Tech Stack

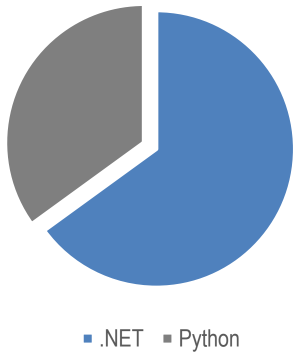


.NET Usage

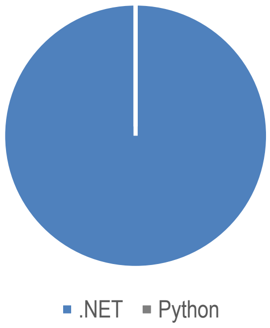
- .NET very powerful at processing many requests and data at speed
- Most software applications written for this system in .NET, many packages
- A lot of packages for Machine Learning, AI, NLP and standard tools have continued to grow in this space
- .NET Core has provided better compatibility and flexibility to use together with Python and other languages.
  - Use the strengths of all languages

Programming Language Usage

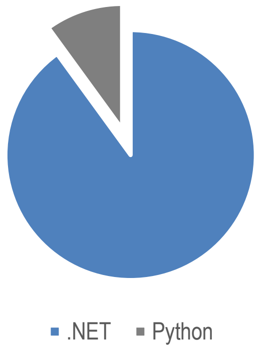
Analytcs / Machine Learning



Data Processing



Software Applications In-house Developed



Questions?