SmarterAnalytics

Oleksandr Romanko, Ph.D.
Senior Research Analyst, Watson Financial Services, IBM
Adjunct Professor, University of Toronto

Workshop on Risk Analytics and Artificial Intelligence in Finance: Hans-On

CASCON Workshop on Risk Analytics and AI in Finance
November 6, 2019

Please note:

IBM Risk Analytics statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

🕮 Notebook 1 – Portfolio Simulation Modeling

Financial portfolio simulation modeling – example 1

- We want to invest $1000 in the US stock market for 1 year: $v_0 = 1000$
- Invest into the S&P 500 market index (index fund)
- Value of investment at the end of year 1: v_1
- Market return over the time period [0,1) is $r_{0,1}$

$$v_1 = v_0 + r_{0,1} \cdot v_0 = (1 + r_{0,1})v_0$$

- Generate scenarios for the market return over the year and compute v_1
  - decide on the number of scenarios and the set of scenarios for $r_{0,1}$
  - generate scenarios
    - ✓ use historic scenarios
    - ✓ draw randomly from historic scenarios (bootstrapping)
    - ✓ draw random numbers from the assumed distribution (Monte Carlo)
  - visualize and analyze the approximate probability distribution of v_1
- In our example we assume that the return of the market over the next year follow Normal distribution

Financial portfolio simulation modeling – example 1

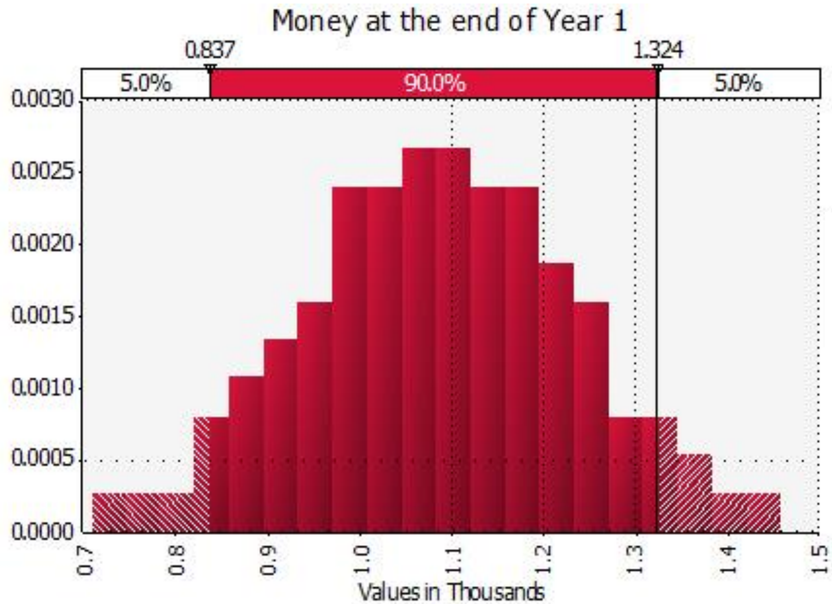
- Between 1977 and 2007, S&P 500 returned 8.79% per year on average with a standard deviation of 14.65%
- Generate 100 scenarios for the market return over the next year (draw 100 random numbers from a Normal distribution with mean 8.79% and standard deviation of 14.65%):

```
r01  = random.normal(0.0879, 0.1465, 100)
```

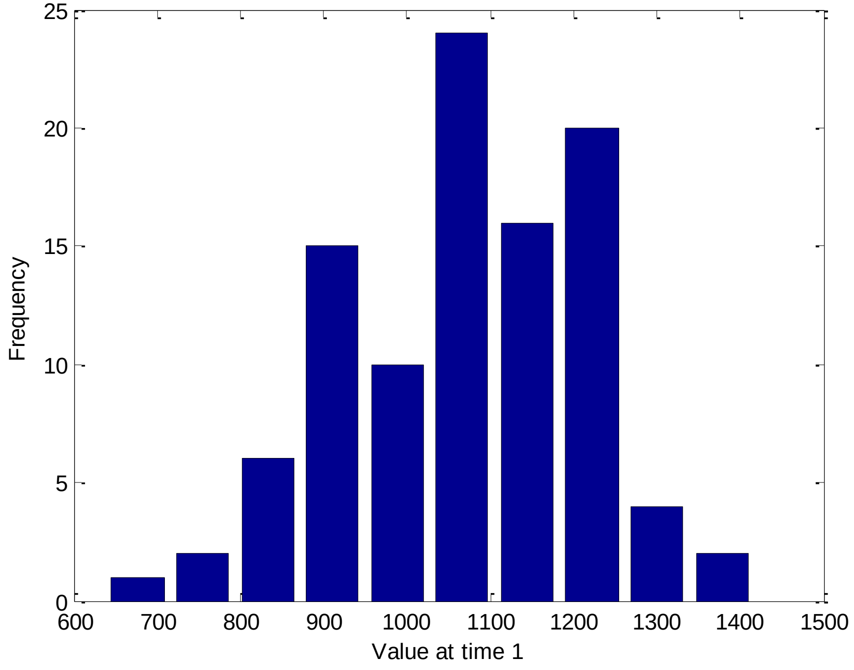
```
 0.099278
-0.004262
 ...
 0.488364
-0.119054
```

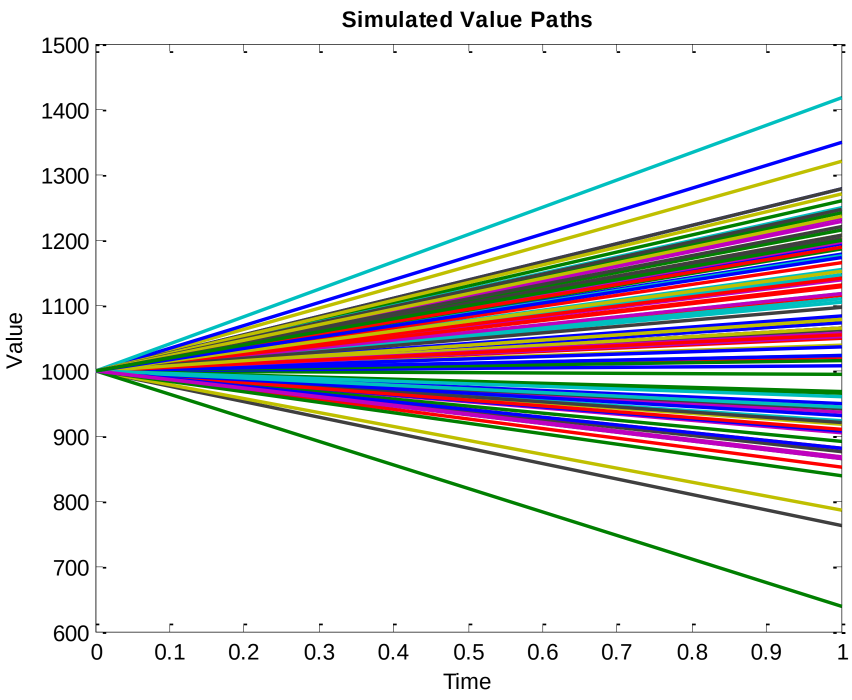
- Compute and plot $v_1 = (1 + r_{0,1})v_0$

Number of values	100
Mean	$ 1,087.90
Std Deviation	$ 146.15
Skewness	0.0034442
Kurtosis	2.871695
Mode	$ 1,118.96
5% Perc	$ 837.40
95% Perc	$ 1,324.00
Minimum	$ 708.81
Maximum	$ 1,458.52



Financial portfolio simulation modeling – example 1





Why use simulation?

- Example 1 illustrates very basic Monte Carlo simulation system
- Simulation allows us to evaluate (approximately) a function of a random variable
  - in example 1 the function is simple $v_1 = (1 + r_{0,1})v_0$
  - given distribution of $r_{0,1}$, in some cases we can compute distribution of v_1 in closed form, e.g., if $r_{0,1}$ followed a Normal distribution, then v_1 also follows a Normal distribution with mean $(1 + \mu_{0,1})v_0$ and standard deviation $\sigma_{0,1}v_0$
  - if $r_{0,1}$ was not Normally distributed, or if the output variable v_1 were a more complex function of the input variable $r_{0,1}$, it would be difficult and practically impossible to derive the probability distribution of v_1 from the probability distribution of $r_{0,1}$
- Other advantages of simulation:
  - simulation enables visualizing probability distribution resulting from compounding probability distributions of multiple input variables (example 2)
  - simulation allows incorporating correlations between input variables (example 3)
  - simulation is a low-cost tool for checking the effect of changing a strategy on an output variable of interest (example 4)
- Next, we extend example 1 to illustrate such situations

Financial portfolio simulation modeling – example 2

- You are planning for retirement and decide to invest in the market for the next 30 years (instead of only the next year as in example 1). Your initial capital is still $v_0 = 1000$
- Assume that every year your investment returns from investing into the S&P 500 will follow a Normal distribution with the mean and standard deviation as in example 1.
- Value of investment after 30 years: v_{30}
- The return over 30 years will depend on the realization of 30 random variables

$$v_{30} = (1 + r_{0,1}) \cdot (1 + r_{1,2}) \cdot \ldots \cdot (1 + r_{29,30}) \cdot v_0$$

$$r_{0,t} = (1 + r_{0,1})(1 + r_{1,2}) \ldots (1 + r_{t-1,t}) - 1$$

$$v_{0,t} = (1 + r_{0,t}) v_0$$

- Observations:
  - sum of Normal random variables is Normal
  - here we have multiplication of Normal random variables, is it Normal?

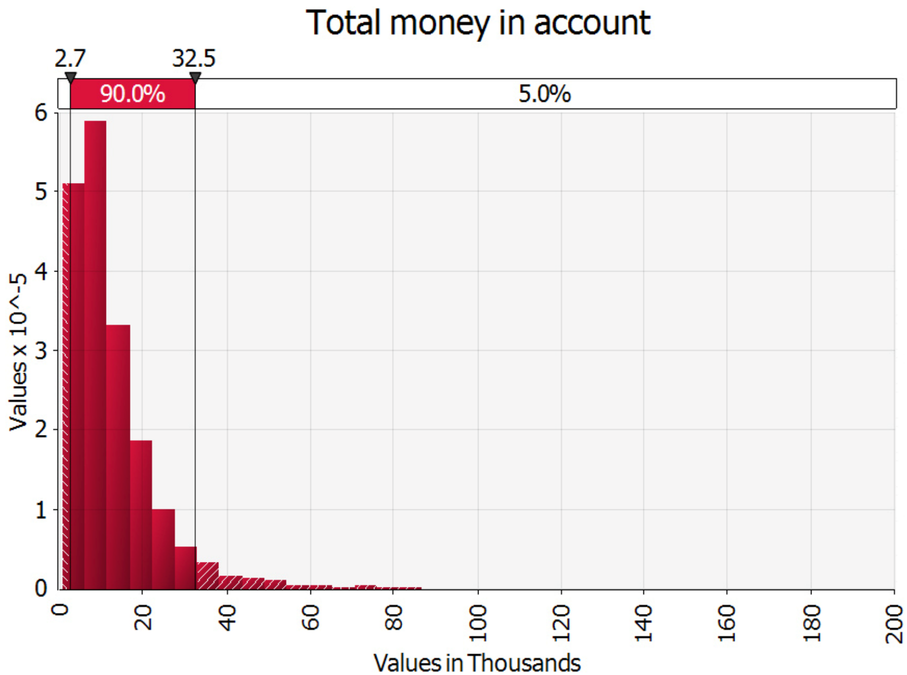
Financial portfolio simulation modeling – example 2

- Between 1977 and 2007, S&P 500 returned 8.79% per year on average with a standard deviation of 14.65%
- Simulate 30 columns of 100 observations each of single period returns:

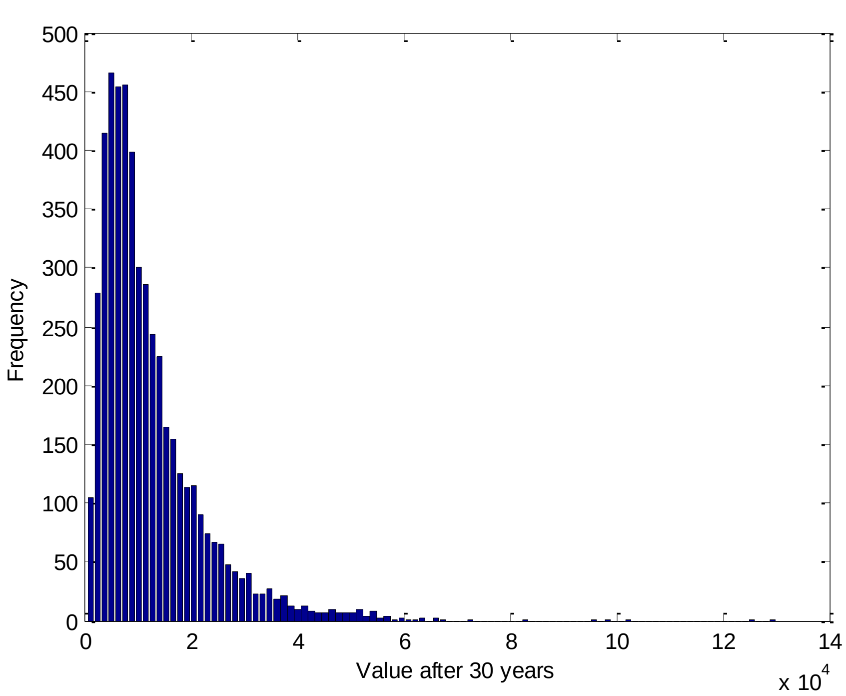
```
r_speriod30  = random.normal(0.0879, 0.1465, (100, 30))
 0.323770    0.188574    ...    0.024316
 0.060499    0.142391    ...    0.093383
 ...         ...         ...    ...
-0.019156   -0.120207    ...    0.071931
 0.289694    0.038724    ...    0.356291
```

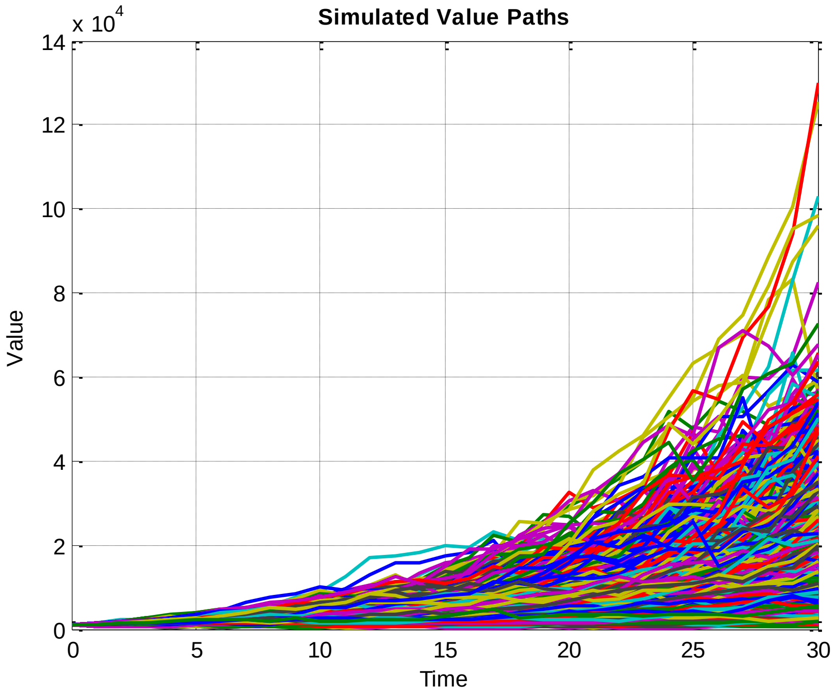
- Compute and plot $v_{30} = (1 + r_{0,30})$

Number of values	5000
Mean	$ 12,587.62
Std Deviation	$ 10,948.39
Skewness	3.349066
Kurtosis	28.24214
Mode	$ 4,458.97
5% Perc	$ 2,655.55
95% Perc	$ 32,481.38
Minimum	$ 609.75
Maximum	$194,355.00



Financial portfolio simulation modeling – example 2





Financial portfolio simulation modeling – example 3

- You are planning for retirement and decide to invest in the market for the next 30 years. Your initial capital is $v_0 = 1000$
- You have an opportunity to invest in stocks and Treasury bonds:
  - allocate 50% of your capital to the stock market (S&P 500 index fund) today
  - allocate 50% of your capital to bonds today
- Assume that every year your investment returns from investing into the S&P 500 and Treasury bonds will follow a Normal distribution with the mean and standard deviation as in example 2 (for S&P 500), mean 4% and standard deviation 7% for bonds. Assume correlation -0.2 between the stock market and the Treasury bond market.
- Covariance matrix:

$$\begin{pmatrix} 0.1465^2 & -0.2 \cdot 0.1465 \cdot 0.07 \\ -0.2 \cdot 0.1465 \cdot 0.07 & 0.07^2 \end{pmatrix} = \begin{pmatrix} 0.0215 & -0.0021 \\ -0.0021 & 0.0049 \end{pmatrix}$$

- Value of investment after 30 years: v_{30}

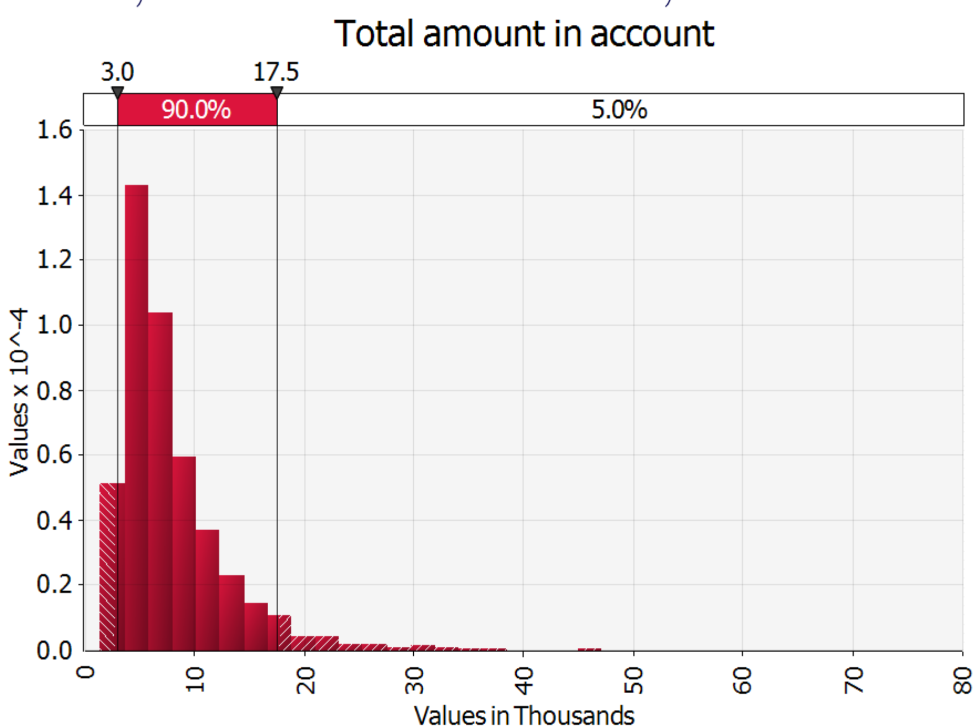
Financial portfolio simulation modeling – example 3

- Simulate 30 years of 100 observations each of single period correlated returns:

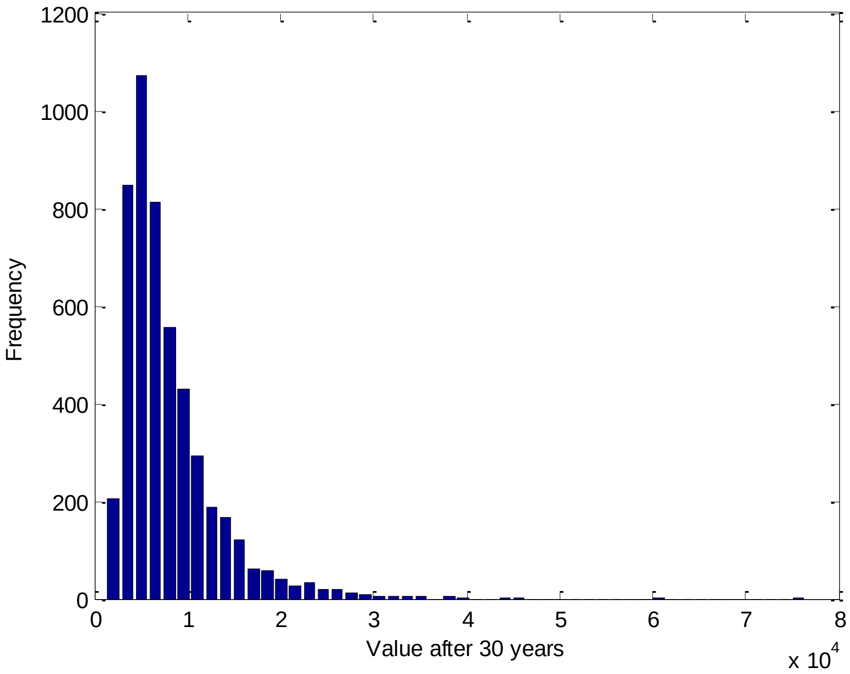
```
scenarios = random.multivariate_normal(mu, covMat, Ns)
for year in range(1, 31):
    scenarios = random.multivariate_normal(mu, covMat, Ns)
    stockRet *= (1 + scenarios[:,0])
    bondsRet *= (1 + scenarios[:,1])
```

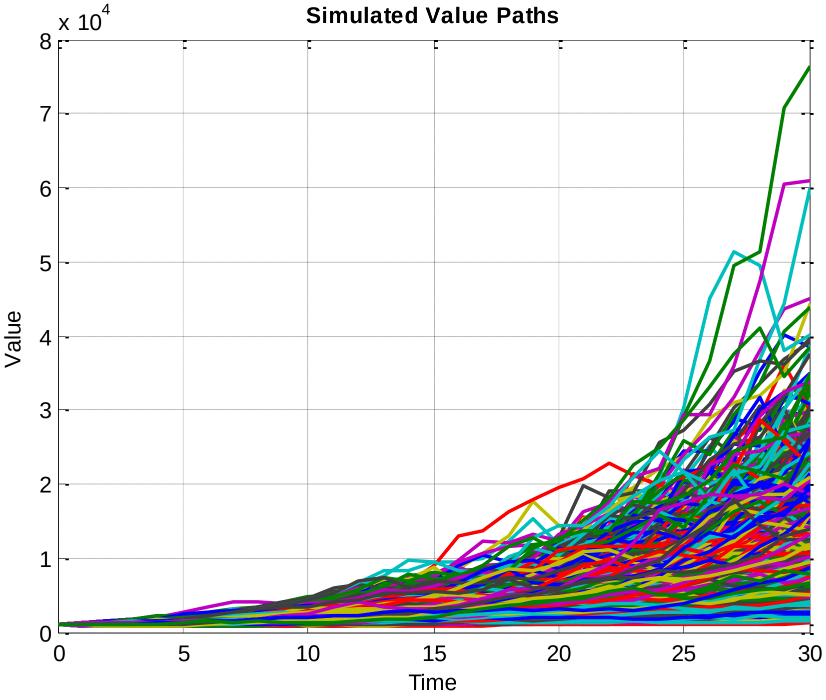
- Compute and plot $v_{30} = 0.5v_0(1 + r^s_{0,30}) + 0.5v_0(1 + r^b_{0,30})$

Number of values	5000
Mean	$ 7,892.80
Std Deviation	$ 5,233.10
Skewness	2.921482
Kurtosis	20.48869
Mode	$ 5,050.96
5% Perc	$ 2,951.82
95% Perc	$17,457.43
Minimum	$ 1,408.63
Maximum	$79,729.34



Financial portfolio simulation modeling – example 3





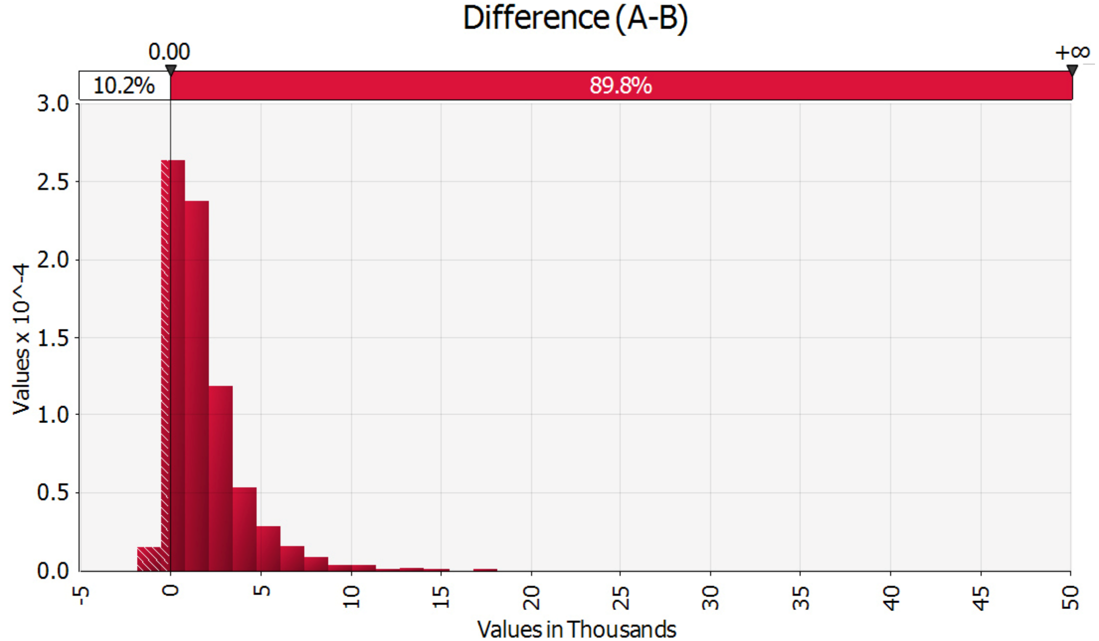
Financial portfolio simulation modeling – example 4

- Using scenario generation procedure from example 3 for decision-making
- Compare portfolios:
  - 50-50 portfolio allocation in stocks and bonds (Strategy A)
  - 30-70 portfolio allocation in stocks and bonds (Strategy B)

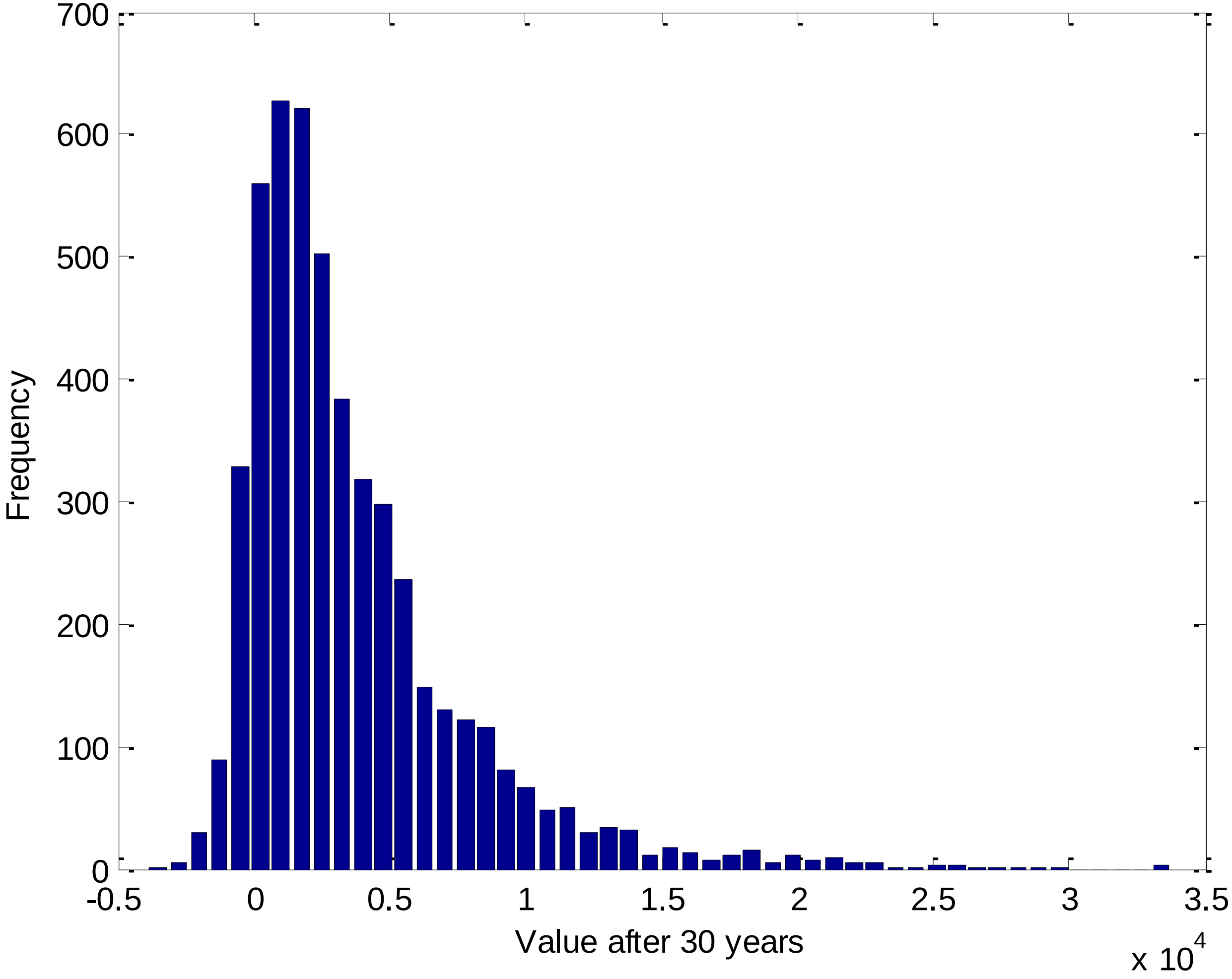
```
v30comp = []
for w in arange(0.2, 1.01, 0.2):
    v30comp += [w * v0 * stockRet + (1 - w) * v0 * bondsRet]
```

- Compute and plot $v_{30} = w_s v_0 (1 + r^s_{0,30}) + w_b v_0 (1 + r^b_{0,30})$

Number of values	5000
Mean	$ 1,865.13
Std Deviation	$ 2,214.87
Skewness	3.506451
Kurtosis	40.18968
Mode	$ 687.75
5% Perc	$ -254.41
95% Perc	$ 6,027.23
Minimum	$-1,829.78
Maximum	$45,972.08



Financial portfolio simulation modeling – example 4

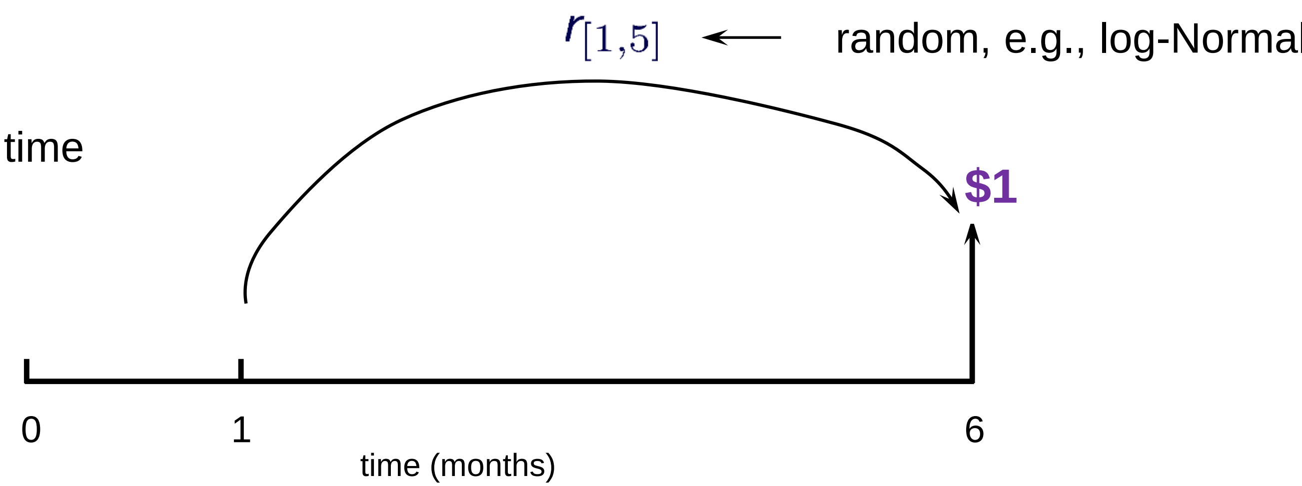


Notebook 2 – Option Valuation

Pricing bond – example

1. Model the risk factor: 5-month interest rate ($r_{[1,5]}$)

$t = 0$ – current time
$\tau = 1$ – risk horizon
$T = 6$ – bond maturity time



2. Compute the zero coupon bond price at time $\tau = 1$

uncertain value (price)

$$v_1 = f(r_{[1,5]}) = e^{-r_{[1,5]} \cdot \frac{5}{12}}$$

sampled value (price)

$$v_{i1} = f(r_{i[1,5]}) = e^{-r_{i[1,5]} \cdot \frac{5}{12}}$$

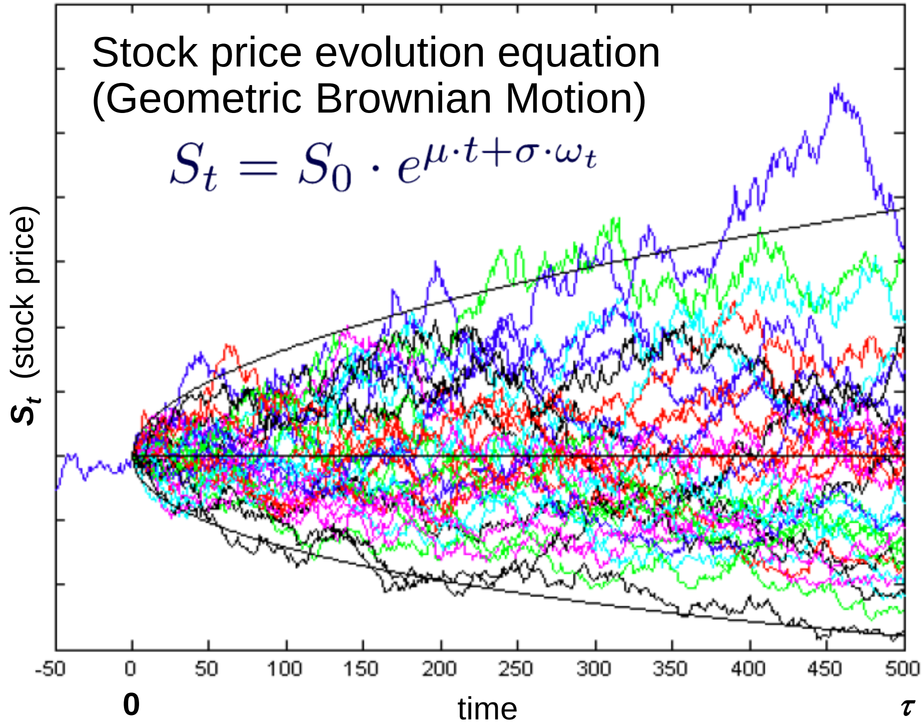
3. Compute the price at $\tau = 1$ if the bond pays a coupon c at time $t = 3$

sampled value (price)

$$v_{i1} = f(r_{i[1,2]}, r_{i[1,5]}) = c \cdot e^{-r_{i[1,2]} \cdot \frac{2}{12}} + e^{-r_{i[1,5]} \cdot \frac{5}{12}}$$

Pricing option – example

1. Model the risk factor: underlying stock price (S_t)



τ is the time at which we compute the option value

T is the option maturity time

2. Model the risk factor: discount rate ($r_{[\tau, T-\tau]}$)

3. Compute option price for path (scenario) i at time τ

$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{i\tau})$$ ← European option, price with Black-Scholes formula

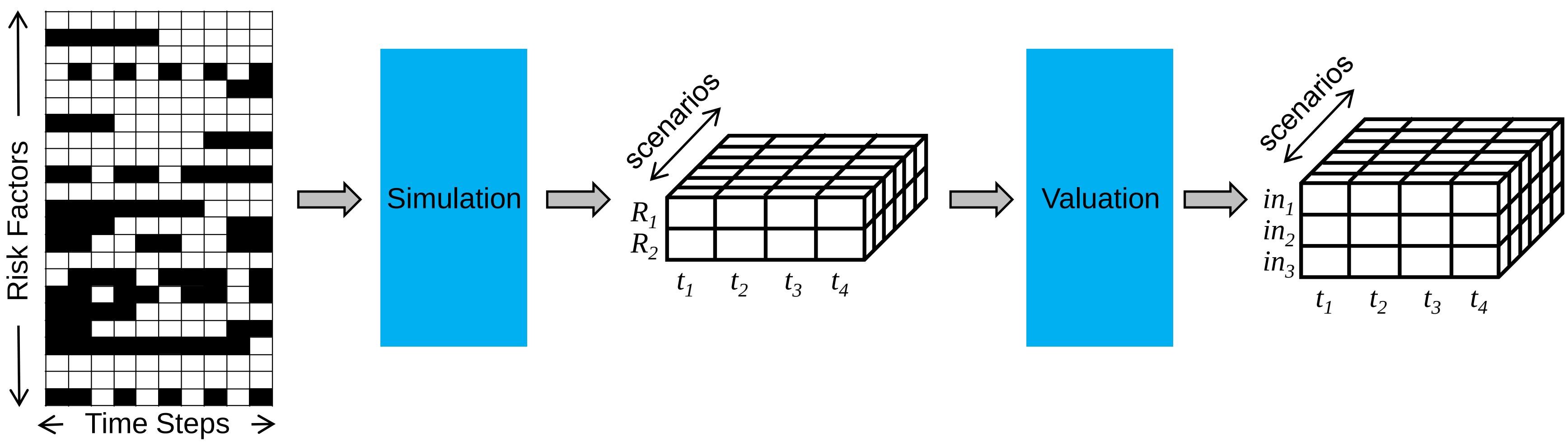
$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{it_1}, S_{it_2}, \ldots, S_{iT})$$ ← Asian option, price is path-dependent

Mark-to-Future framework

Historical Time Series
- Interest rates
- Exchange rates
- Equity indexes
- Commodities
- Individual securities

Risk Factor Scenarios
- Consistent paths of risk factor changes
- Financial instruments may be path-dependent

Instrument Scenarios
- Consistent paths of instrument values
- Mark-to-Future cube



Simulation Methods
- Historical sampling
- Stochastic models
- Calibration
- Codependence
- Transformations (PCA)

Valuation Methods

$$v_{ijt} = f_j(\boldsymbol{R}_{i,t\leq\tau})$$

- Linear factor models
- Analytic
- Numerical methods

Portfolio valuation

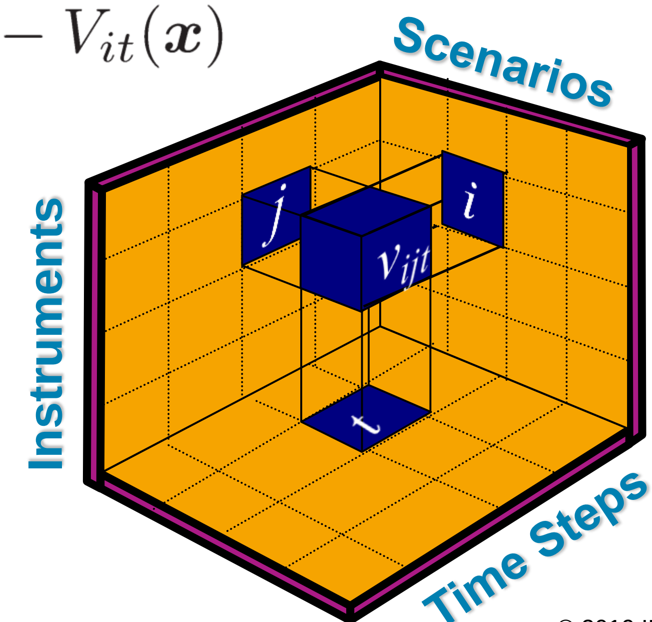
- A portfolio is a set of positions, $\boldsymbol{x}$, where x_j is the number of units of instrument j, $j = 1, \ldots, N$
- From the instrument values in the MtF cube, the portfolio value at time t in scenario i is

$$V_{it}(\boldsymbol{x}) = \sum_{j=1}^{N} v_{ijt} \cdot x_j$$

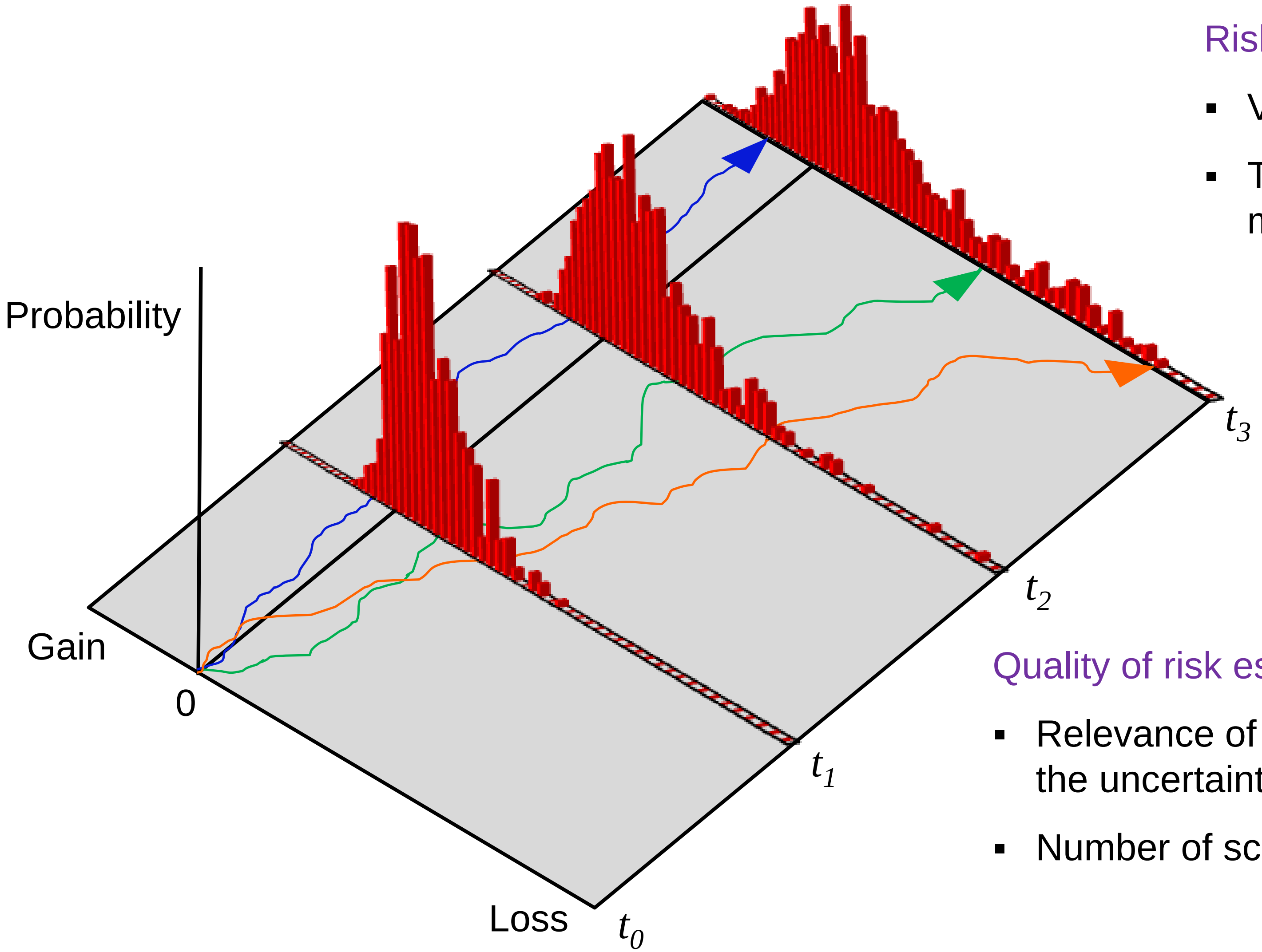
- Given the initial portfolio value, it is straightforward to compute changes in value (profits and losses) and returns in each scenario

$$\text{loss} \longrightarrow \quad \ell_{it}(\boldsymbol{x}) = V_0(\boldsymbol{x}) - V_{it}(\boldsymbol{x})$$

- An MtF cube can be used to value portfolios that hold any subset of its instruments
  - No need to re-simulate when positions change
  - $\boldsymbol{x}$ are decision variables in portfolio optimization



Empirical portfolio loss distributions over time



Risk

- Variance
- Tail-based risk measures

Quality of risk estimates depends on

- Relevance of scenarios (capture the uncertainty)
- Number of scenarios

Modeling asset price dynamics

- Financial time series is a sequence of observations of the values of a financial variable over time:
  - asset price (index level) or asset value
  - asset (index) return
  - interest rate
  - exchange rate

S&P 500 index level between August 19, 2005 and August 19, 2009 (weekly observations)

Modeling asset price dynamics

- Most important **time series** properties are drift and volatility:
  - **drift** – direction of any observable trend in the time series (positive drift from August 2005 until middle of 2007, negative drift from middle of 2007 until the beginning of 2009)
  - **volatility** – measure of variation over time, i.e., standard deviation or variance (volatility was smaller from August 2005 until middle of 2007 and increased dramatically between middle of 2007 and beginning of 2009)
  - volatility to drift (volatility increases when price level increases, volatility decreases when price level increases, volatility remains constant independently of the current price level)
  - continuity – is the time series smooth or there are jumps (quite smooth from August 2005 until middle of 2007 and dramatic drops in price levels between middle of 2007 and beginning of 2009, especially in the fall of 2008)

Notation

- S_t value of the underlying variable (price, interest rate, index level, etc.) at time *t*

- S_{t+1} value of the underlying variable (price, interest rate, index level, etc.) at time *t*+1

- w_t random error term observed at time *t* (it follows random Normal distribution $\mathcal{N}(0, \sigma)$)

- ε_t realization of a standard Normal random variable $\mathcal{N}(0, 1)$ at time *t* (for modelling random error term)

Geometric random walks

- Model return as

$$r_t = \mu + \sigma \cdot \epsilon_t \qquad\qquad r_t = \frac{S_{t+1} - S_t}{S_t}$$

- $\epsilon_0, \ldots, \epsilon_t$ is a sequence of independent standard Normal random variables
- Returns are normally distributed, return over each time interval of length 1 has mean μ and standard deviation σ
- If we know price at time *t*, we can compute price at time *t*+1 as

$$\begin{aligned} S_{t+1} &= S_t \cdot \frac{S_{t+1}}{S_t} = S_t \cdot \left(\frac{S_t}{S_t} + \frac{S_{t+1} - S_t}{S_t} \right) \\ &= S_t \cdot \left(1 + \frac{S_{t+1} - S_t}{S_t} \right) = S_t \cdot (1 + r_t) \\ &= S_t + \mu \cdot S_t + \sigma \cdot S_t \cdot \epsilon_t \end{aligned}$$

Geometric random walks

- Express price at time *t* in terms of known initial price at time 0

$$S_t = S_0 \cdot \frac{S_1}{S_0} \cdot \ldots \cdot \frac{S_{t-1}}{S_{t-2}} \cdot \frac{S_t}{S_{t-1}} = S_0 \cdot (1 + r_0) \cdot \ldots \cdot (1 + r_{t-1})$$

- Problem: product of Normal random variables is not a Normal random variable
- Solution: take logarithms of both sides of the equation and note that

$$\ln(1 + r_t) = \ln(1 + \frac{S_{t+1} - S_t}{S_t}) = \ln(\frac{S_{t+1}}{S_t}) = \ln(S_{t+1}) - \ln(S_t)$$

- Assume that log returns (not returns) are independent and follow Normal distribution with mean μ and standard deviation σ:

$$\ln(1 + r_t) = \ln(S_{t+1}) - \ln(S_t) = \mu + \sigma \cdot \epsilon_t$$

- Now, $\ln(S_t)$ is also Normally distributed (S_t is lognormal random variable)

$$\ln(S_t) = \ln(S_0) + \ln(1 + r_0) + \ldots + \ln(1 + r_{t-1})$$

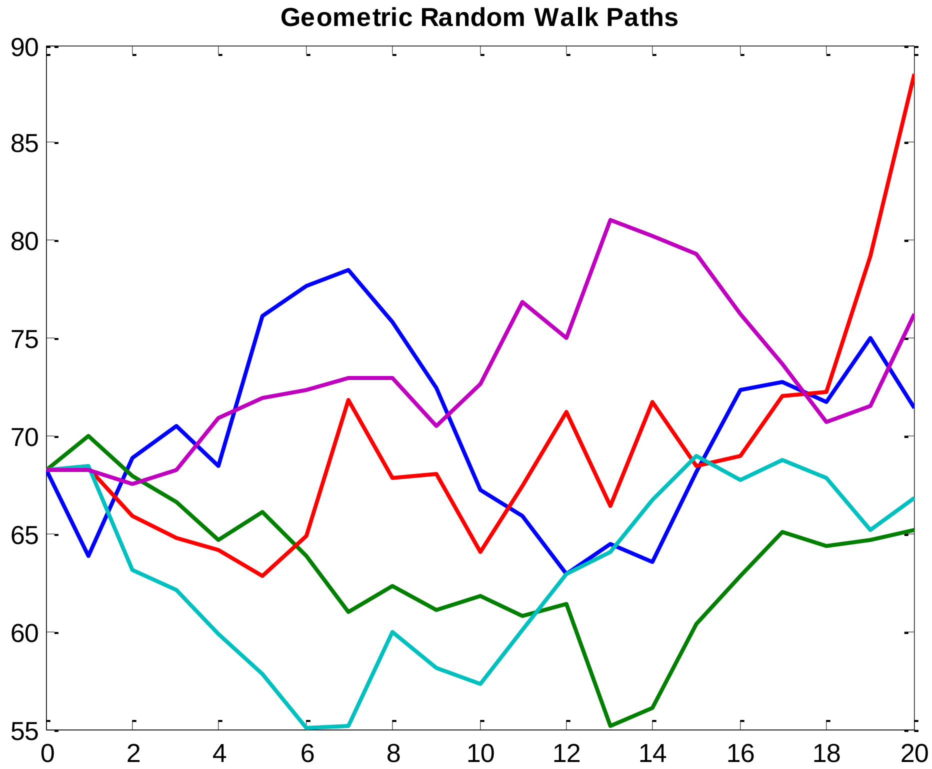
- We can compute closed-form expression for price

$$\ln(S_t) = \ln(S_0) + (\mu - \frac{1}{2} \cdot \sigma^2) \cdot t + \sigma \cdot \sqrt{t} \cdot \epsilon$$

$$S_t = S_0 \cdot e^{(\mu - \frac{1}{2} \cdot \sigma^2) \cdot t + \sigma \cdot \sqrt{t} \cdot \epsilon}$$

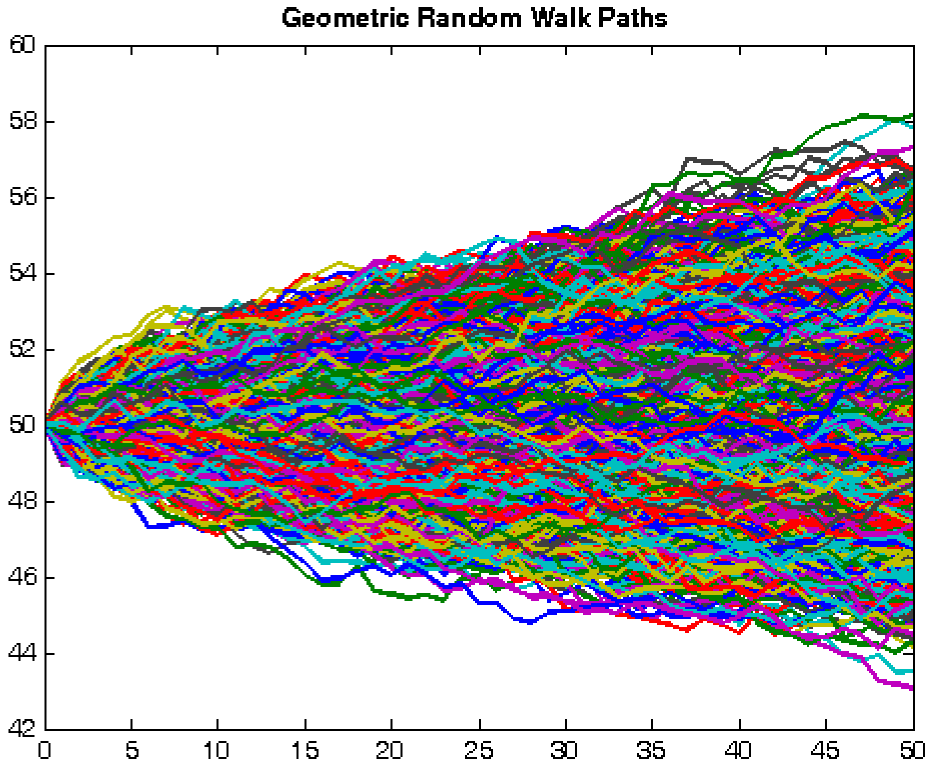
Geometric random walks – example in Python

$$S_{t+1} = S_t \cdot e^{(\mu - \frac{1}{2} \cdot \sigma^2) + \sigma \cdot \epsilon_t}$$



Geometric random walks – example in Python

$$S_{t+1} = S_t \cdot e^{(\mu - \frac{1}{2} \cdot \sigma^2) + \sigma \cdot \epsilon_t}$$



Pricing European option

- Generate a specified number of simulated asset paths and then use those paths to price a standard European Put and Call option. The payoff of the options is given by:

$$P_{\text{call}}(T) = \max(S(T) - K,\ 0)$$

$$P_{\text{put}}(T) = \max(K - S(T),\ 0)$$

  - *S(T)* is the price of the underlying asset at maturity
  - *K* is the strike

Pricing Asian option

- An Asian option is an example of an option that has a path dependent payoff. This makes it ideally suited for pricing using the Monte-Carlo approach.

- Generate a specified number of simulated asset paths and then use those paths to price a standard Asian Put and Call option. The payoff of the options is given by:

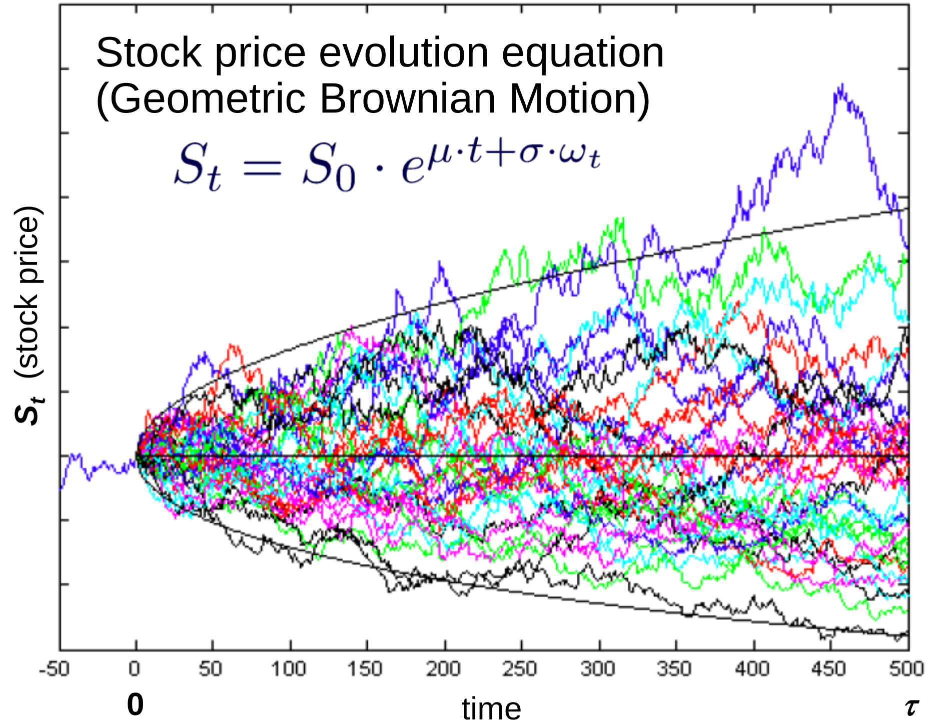
$$P_{\mathrm{call}}(T) = \max(A(0,T) - K,\ 0)$$

$$P_{\mathrm{put}}(T) = \max(K - A(0,T),\ 0)$$

  - $A(0,T)$ is the average price of the underlying asset over the life of the option
  - K is the strike

Pricing option – example

1. Model the risk factor: underlying stock price (S_t)



τ is the time at which we compute the option value

T is the option maturity time

2. Model the risk factor: discount rate ($r_{[\tau, T-\tau]}$)

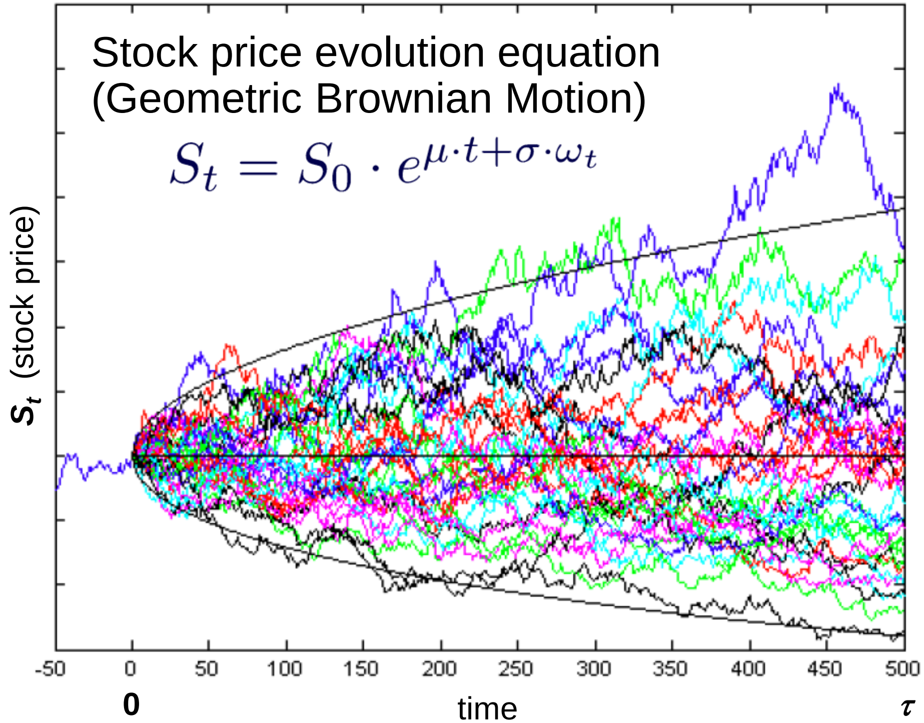
3. Compute option price for path (scenario) i at time τ

$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{i\tau})$$ ⟵ European option

$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{it_1}, S_{it_2}, \ldots, S_{iT})$$ ⟵ Asian option, price is path-dependent

Pricing option – example

1. Model the risk factor: underlying stock price (S_t)



τ is the time at which we compute the option value

T is the option maturity time

2. Model the risk factor: discount rate ($r_{[\tau, T-\tau]}$)

3. Compute option price for path (scenario) $\boldsymbol{i}$ at time $\boldsymbol{\tau}$

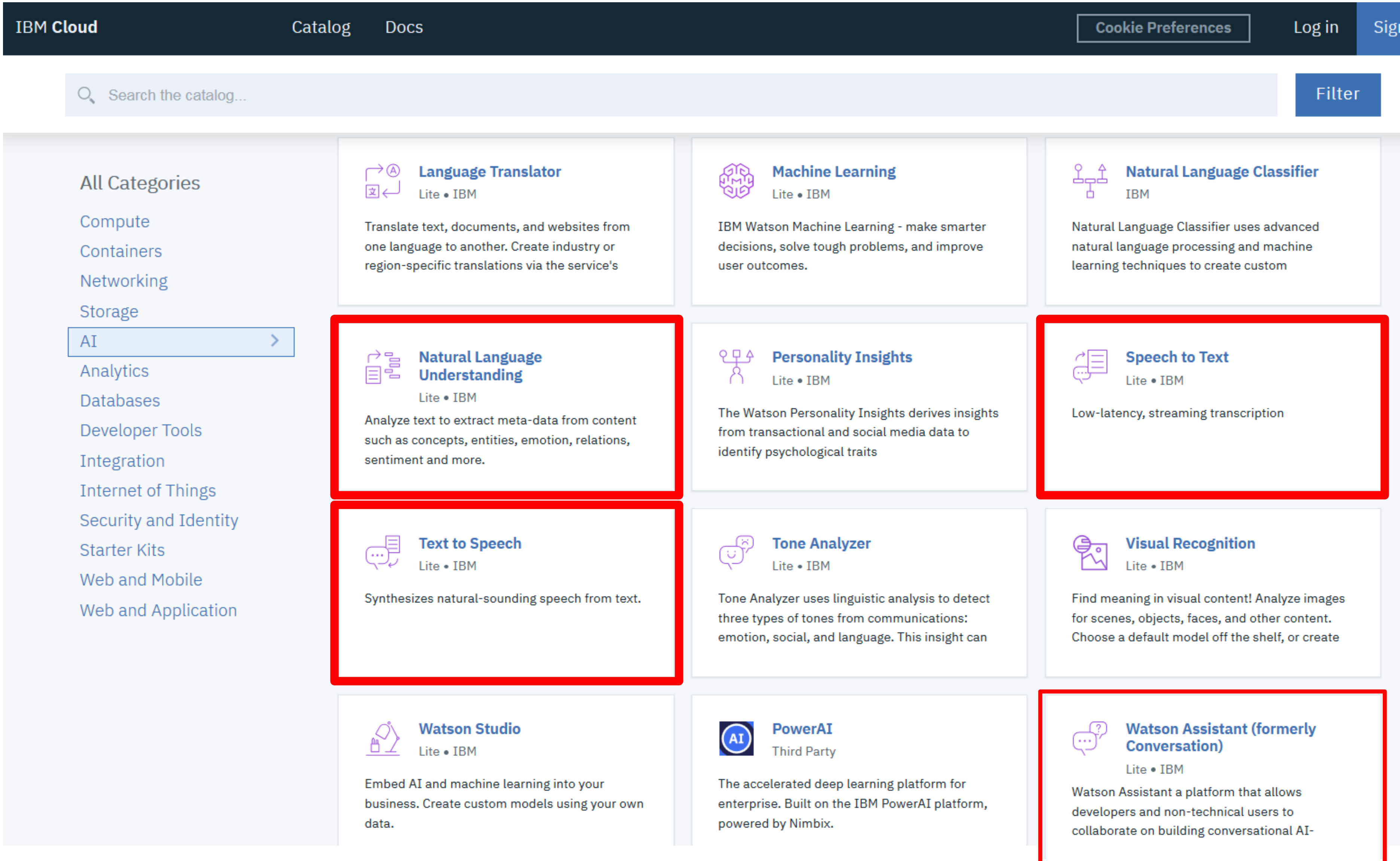
$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{i\tau}) = \max(S_{iT} - K, 0) \cdot e^{r_{i[\tau,T-\tau]} \cdot (T-\tau)}$$ ⟵ European option

$$v_{i\tau} = f(r_{i[\tau,T-\tau]}, S_{it_1}, S_{it_2}, \ldots, S_{iT})$$ ⟵ Asian option, price is path-dependent

$$= \max(\mathrm{mean}(S_{it_1}, S_{it_2}, \ldots, S_{iT}) - K, 0) \cdot e^{r_{i[\tau,T-\tau]} \cdot (T-\tau)}$$

🕮 Notebook 3 – Speech-to-Text, Natural Language Understanding and Text-to-Speech

Watson Services on IBM Cloud



Watson Speech-to-Text, Natural Language Understanding and Text-to-Speech

```
import IPython
IPython.display.Audio("sample.wav")
```

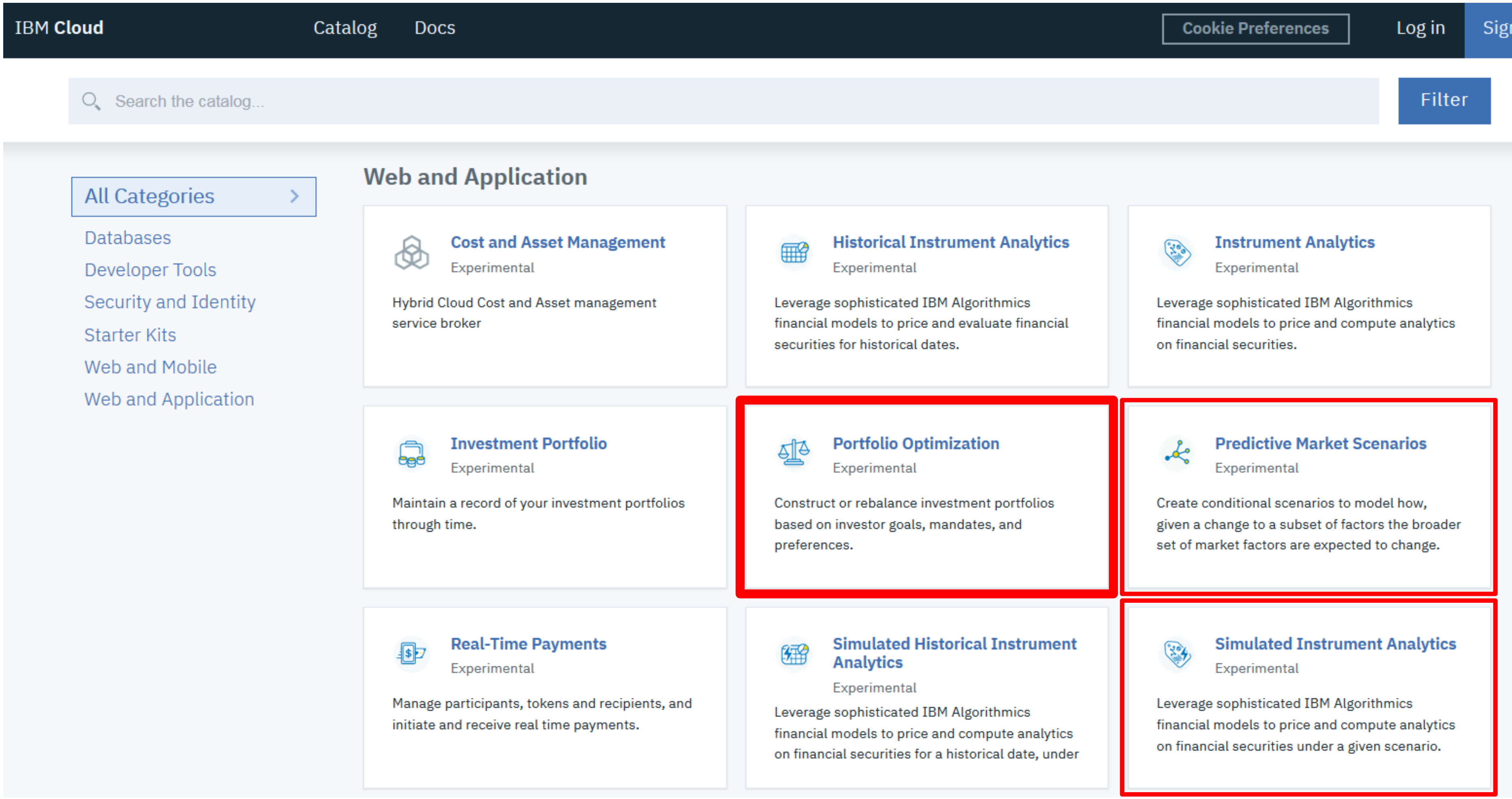
Speech-to-Text

```
from watson_developer_cloud import SpeechToTextV1

# Instantiate the service using your credentials
service = SpeechToTextV1(
    username='72784511-6fe0-44bf-90e2-484acea841c8',
    password='a454oGGRDEWv'
)

audio_file = open("sample.wav", "rb")

result = service.recognize(audio_file, content_type="audio/wav")

result
```

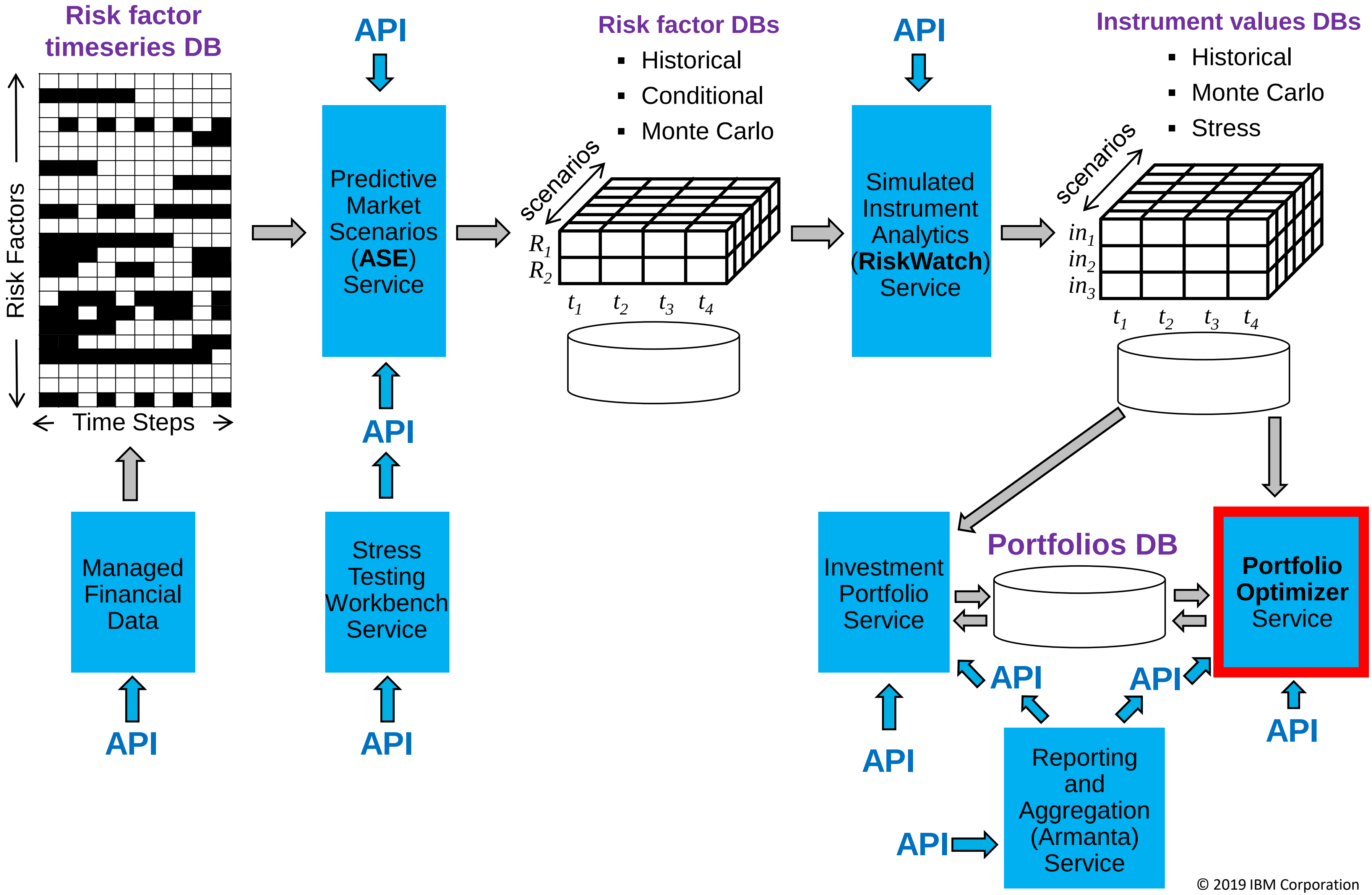
```
{'result_index': 0,
 'results': [{'alternatives': [{'confidence': 0.994,
     'transcript': 'thunderstorms could produce large hail isolated tornadoes and heavy rain '}],
   'final': True}]}
```

Notebook 4 – Portfolio Optimization

Financial Analytics services on IBM Cloud



IBM Cloud services – portfolio optimization



Objective function and constraints

```
"objectives": [
 {
   "sense":     "minimize",
   "measure":   "variance",
   "attribute": "return",
   "portfolio": "Universe",
   "TargetPortfolio": "Aggressive",
   "timestep":  30,
   "description": "minimize tracking error squared (variance of the difference between
                   Universe portfolio and Aggressive benchmark returns) at time 30 days"
 }
],
"constraints": [
 {
   "attribute":   "weight",
   "portfolio":   "HighEnvironmental",
   "InPortfolio": "Universe",
   "relation":    "greater-or-equal"
   "constant":    0.5,
   "description": "Creating an average Environmental score of High"
 },
 {
   "attribute": "weight",
   "members":   "Has Tobacco",
   "relation":  "equal",
   "constant":  0.0,
   "description": "Excluding all securities which have the property Has Tobacco"
 }
]
```

🕮 Notebook 5 – Cognitive Portfolio Optimization

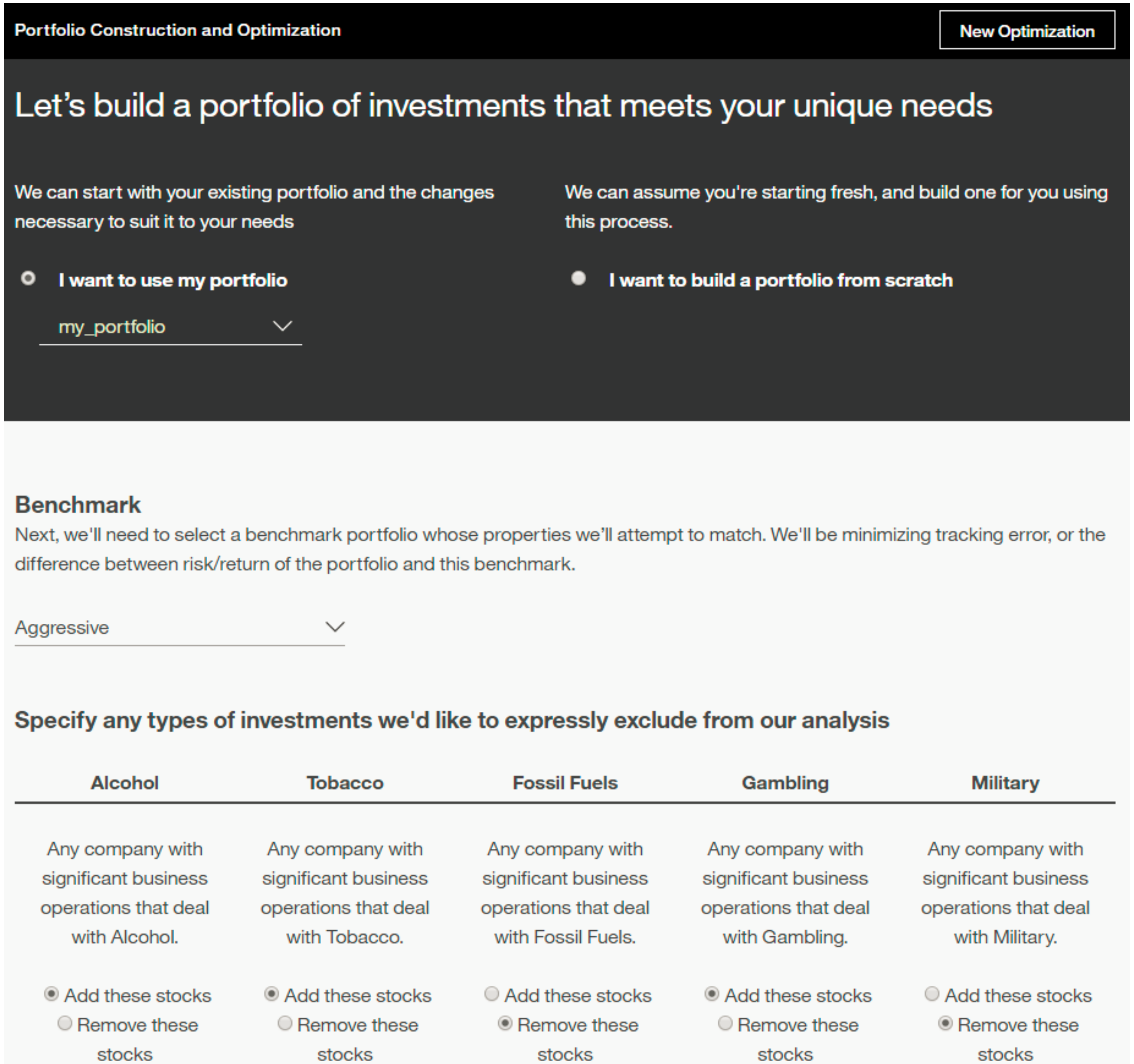
IBM Algo Optimizer interfaces 1 – JSON API

- IBM Algo Optimizer is built with the IBM ILOG CPLEX® library
- Interface allows users to construct optimization problems for financial risk management
- Problem can reference
  - Various financial attributes, e.g., value, return, cash flow, duration, beta, etc.
  - Multiple scenario sets
- Measures include
  - Expectation
  - Variance
  - Linear or quadratic deviations from a target
  - Expected shortfall (CVaR)
- Possible to specify
  - Cardinality restrictions
  - Trading costs
  - Soft constraints
  - Multiple objectives

```
{
   "measure": "expectation",
   "attribute": "return",
   "timestep": 30,
   "relation": "greater-or-equal",
   "constant": 0.0075,
   "description": "expected return of the
                   portfolio is at least 0.75%"
}
```

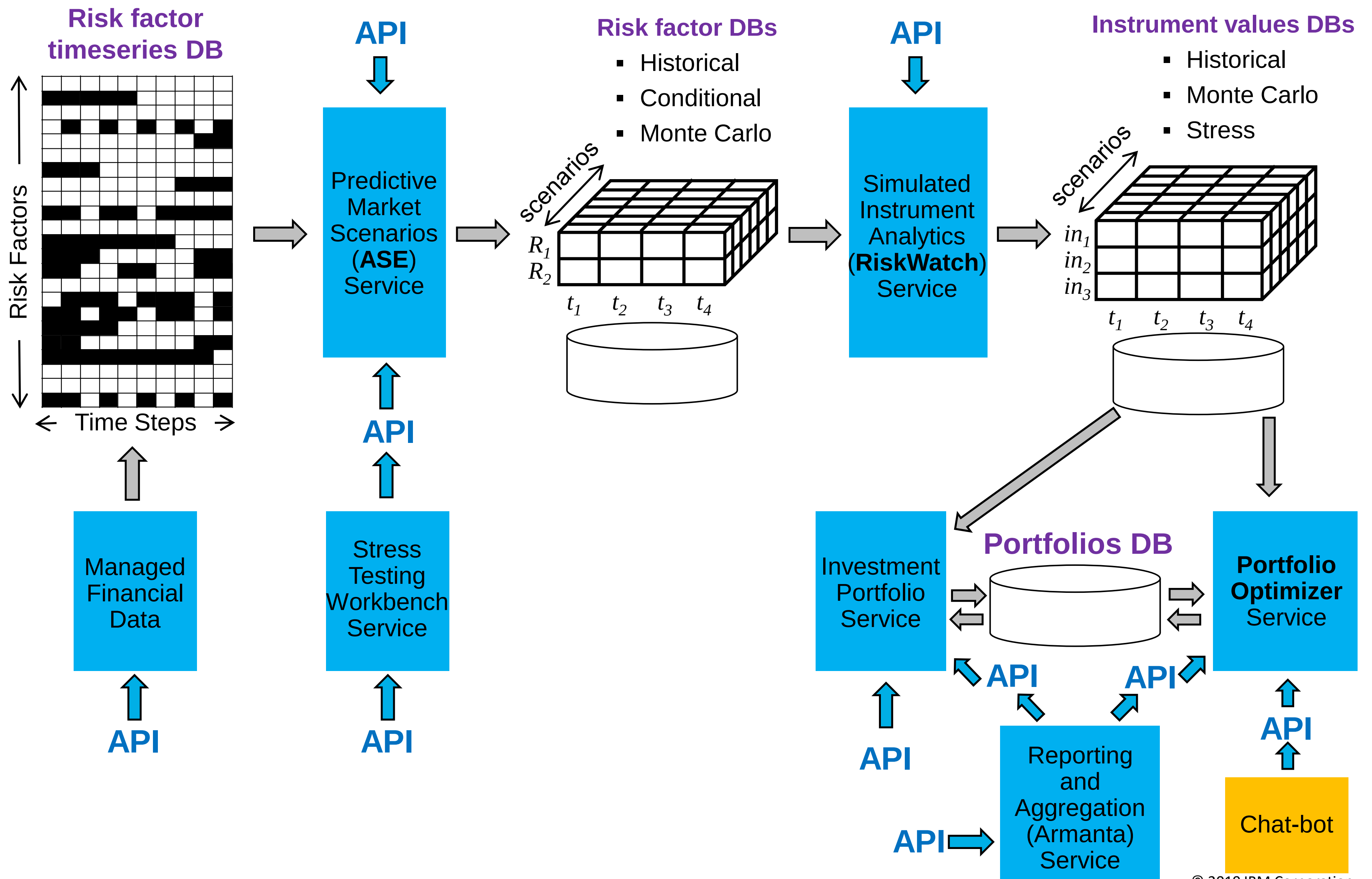
IBM Algo Optimizer interfaces 2 – GUIs

https://developer.ibm.com/code/journey/construct-a-socially-responsible-investment-portfolio/

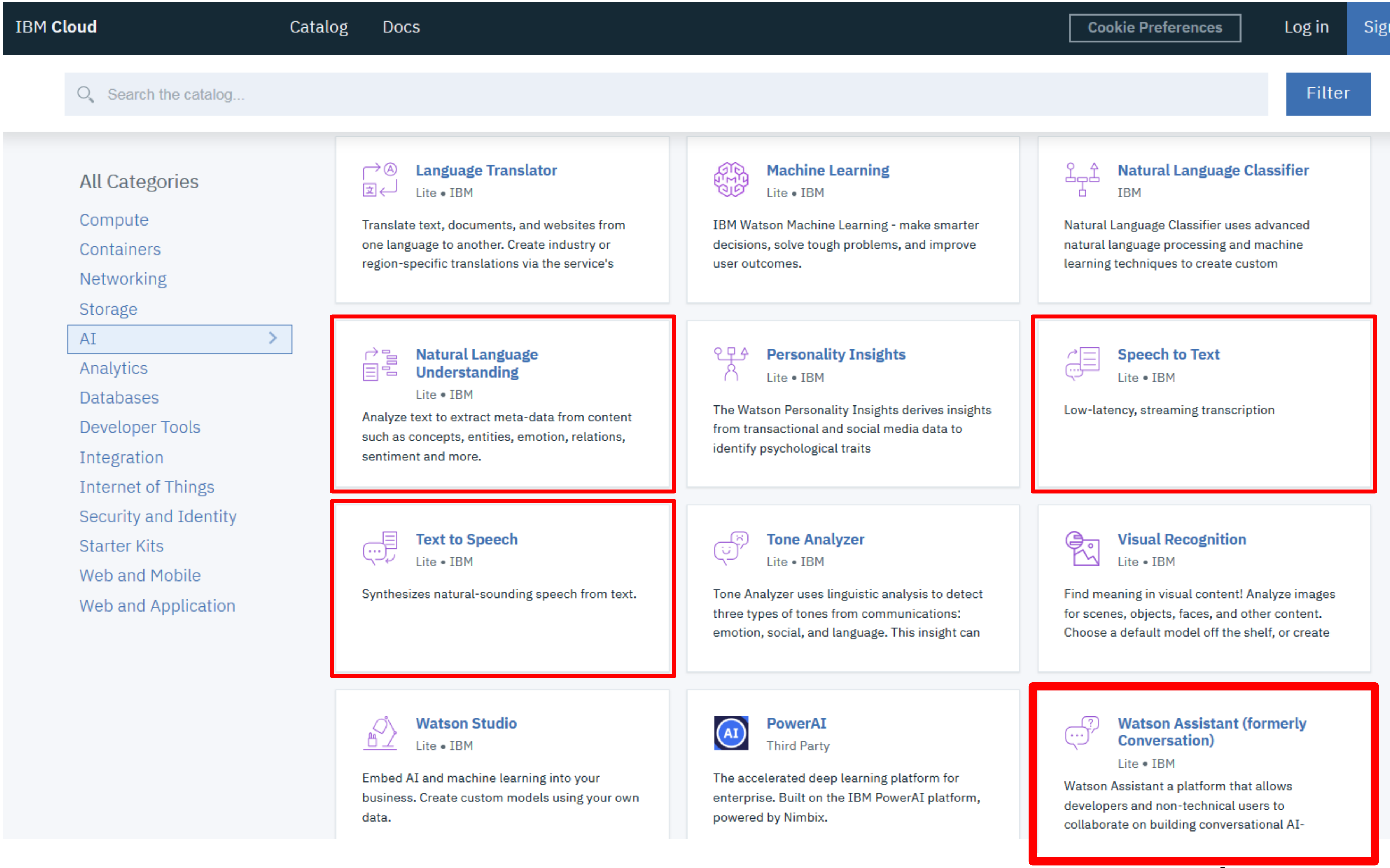


IBM Algo Optimizer interfaces 3 – cognitive and AI

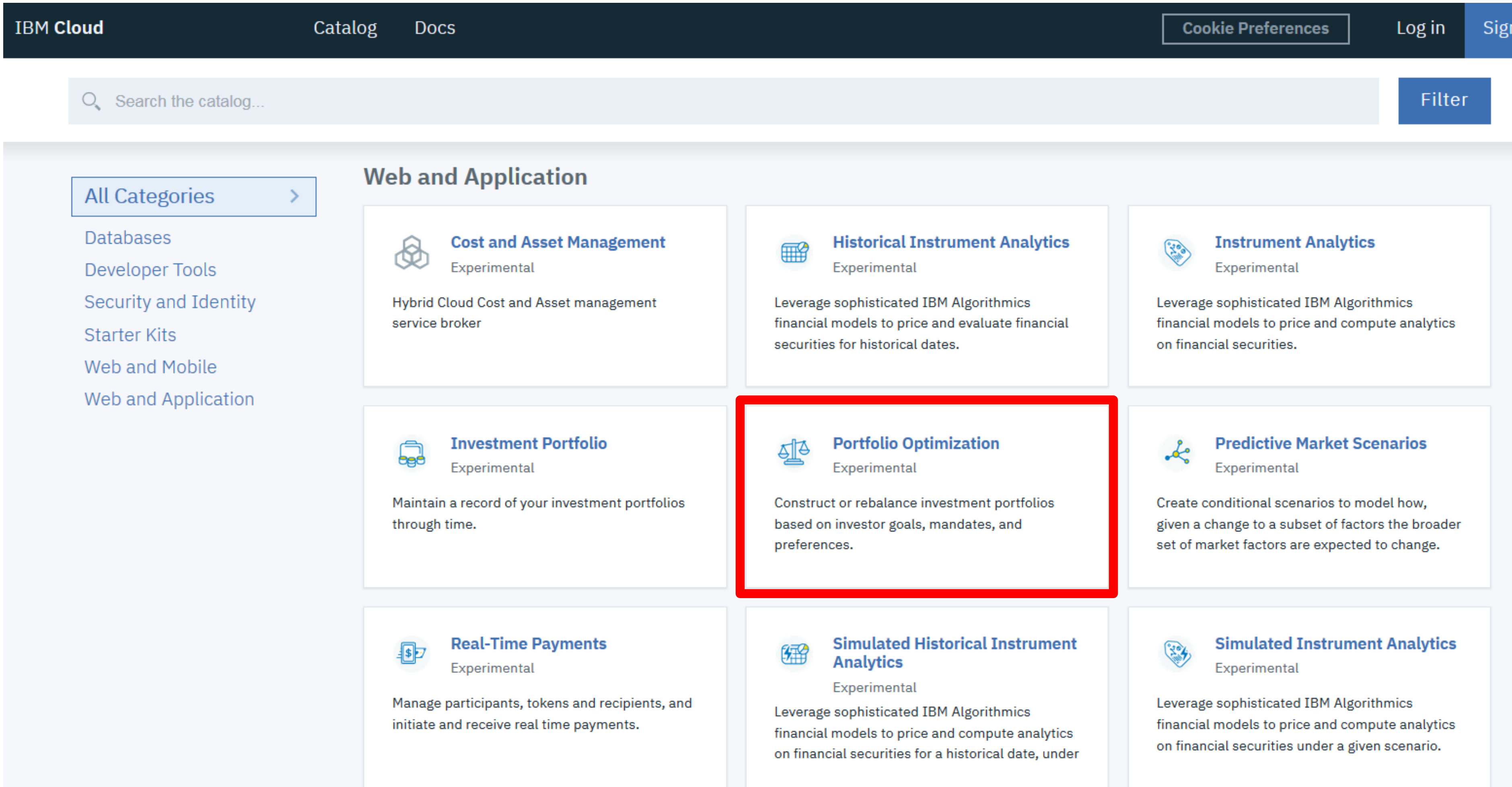
IBM Cloud services – cognitive portfolio selection



Watson Services on IBM Cloud



Portfolio Optimization service on IBM Cloud



Cognitive Portfolio Optimization Service: Watson NLU Service

Watson Natural Language Understanding Service

investor preferences

My name is John. I am **34**. I **work** in **information technology** sector. I would like to invest **$30000** in my **RRSP** this year. I would **like highly diversified portfolio**. I **do not like** having **military** and **dirty assets** in my portfolio. I **care** a lot about **sustainable development**.

Watson Natural Language Understanding:

- ❑ Investing for long-term (34 years, RRSP) -> **aggressive benchmark (stocks)**
- ❑ **Cash inflow of $30000** into current portfolio
- ❑ Highly diversified portfolio -> **weight of each asset <= 5% constraint**
- ❑ No military or dirty assets -> **weight of military and dirty assets = 0 constraint**
- ❑ Care about sustainable development -> **weight of sustainability assets >= 50%**
- ❑ Work in information technology -> **weight of IT assets between 10% and 30%**

Chat-bot for portfolio selection

Methodology

Watson Assistant, Natural Language Understanding

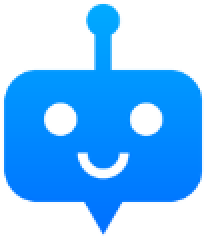
Changing Approach: Converting problem into Chatbot

My name is John. I am 34 and want to invest for 10 years. I work in information technology sector so I want the weight of that sector to be greater than 20%. I would like to invest $30000 in my RRSP this year. I would like a highly diversified portfolio. I do not like having military and tobacco assets in my optimized portfolio. I care a lot about sustainable development.

My name is John.

I am 34 and want to invest for 10 years.

I work in information technology sector so I want the weight of that sector to be greater than 20%.



Chat-bot for portfolio selection

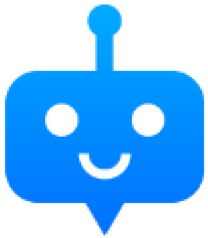
Methodology

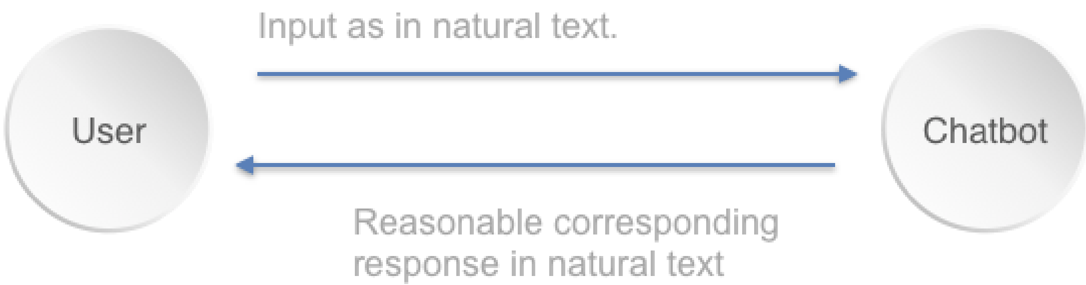
Watson Assistant,
Natural Language
Understanding

My name is John.

I am 34 and want to invest for 10 years.

I work in information technology sector so I want the weight of that sector to be greater than 20%.





Specialized chatbot is a relatively **easy to handle** Natural Language Processing task due to

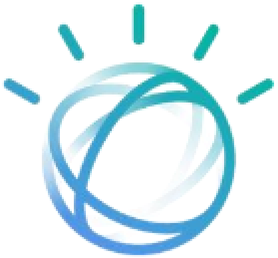
- usually the user's input contains **a certain intent**
- entities and response is relatively fixed

Chat-bot for portfolio selection

Methodology

Watson Assistant, Natural Language Understanding

USER



**Watson Assistant,
Natural Language Understanding**

Portfolio Optimization

I want to use an aggressive benchmark in the optimization process.

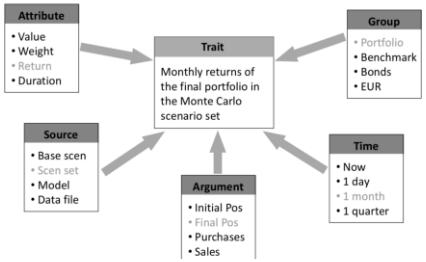
Natural Language Input

Intents Classifier
(Neural Networks Model)

benchmark_selection

Entities Matching

@ benchmark: aggressive



{JSON}

Structured Optimization Request

Chat-bot for portfolio selection

Methodology

Watson Assistant
Intents Classifier
(Neural Networks Model)

Watson Assistant: Intents Classifier

Using **neural networks model** to **classify** the natural text input as different intents or irrelevant. Manually set examples are provided along with the name of intent.

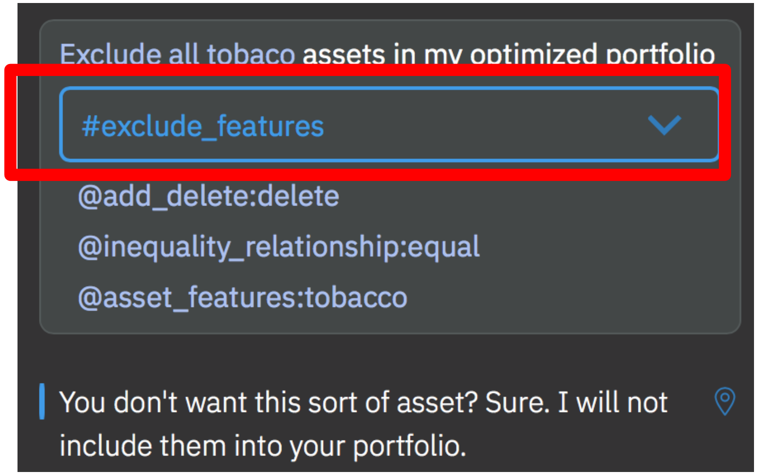
Currently the prototype has 18 intents:

Constraints:
#allow_short_sell. #cash_infusion. #constrain_asset_class. #constrain_geography. #constrain_risk_score. #exclude_features. #portfolio_diversity. #social_responsibility.
Objectives:
#benchmark_selection. #objectives_stating.
Holdings:
#edit_entities..
Chit-chat:
#goodbyes. #greetings. #help.
Functions:
#feedback. #start_optimization. #summarize_constraints.

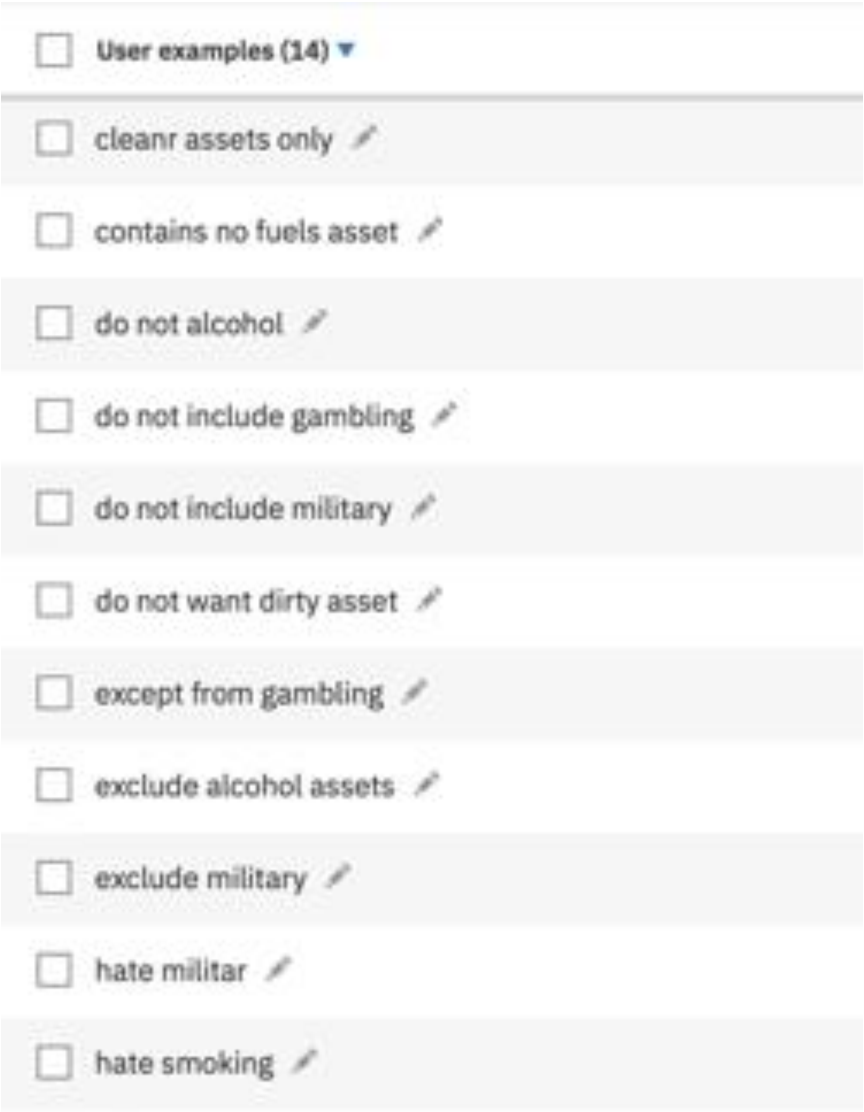
Chat-bot for portfolio selection

Methodology

Watson Assistant
Intents Classifier
(Neural Networks Model)



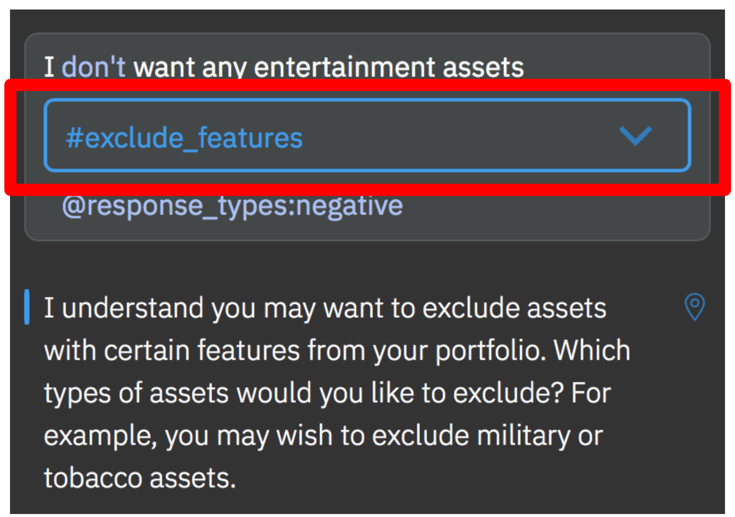
Examples of Intents Classifier



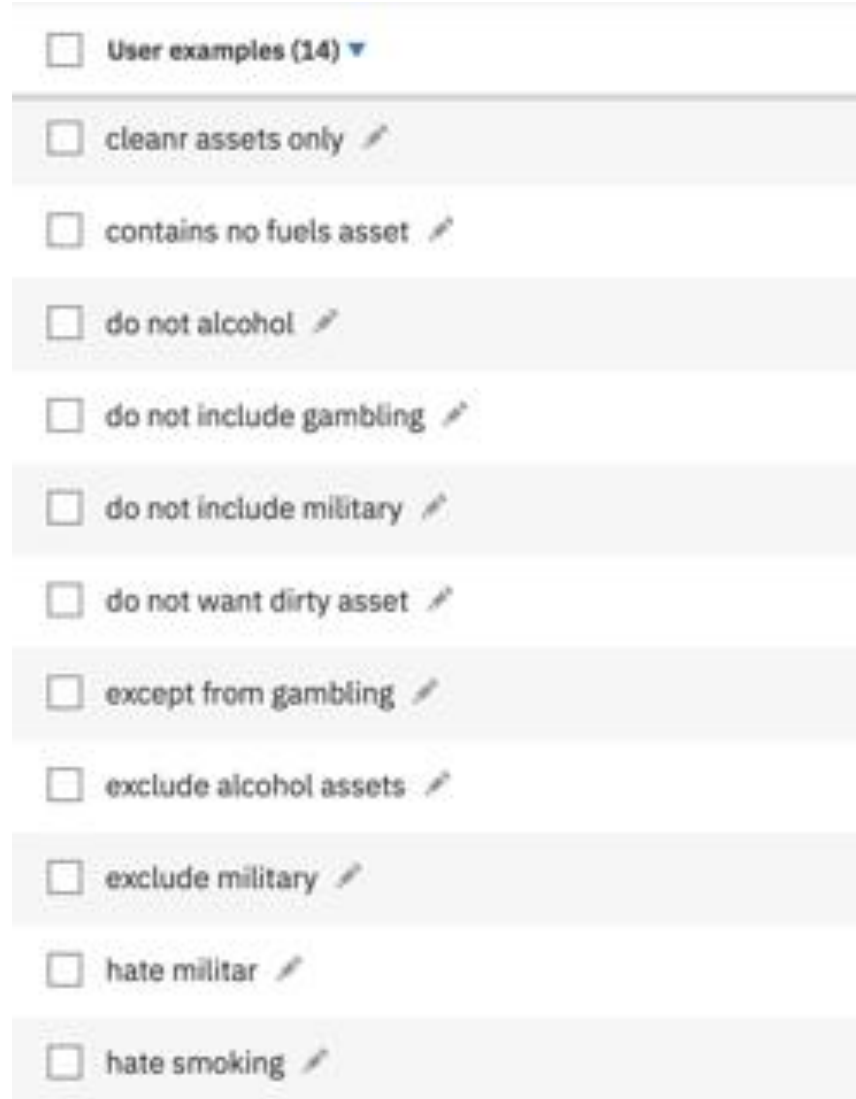
Chat-bot for portfolio selection

Methodology

Watson Assistant
Intents Classifier
(Neural Networks Model)



Examples of Intents Classifier

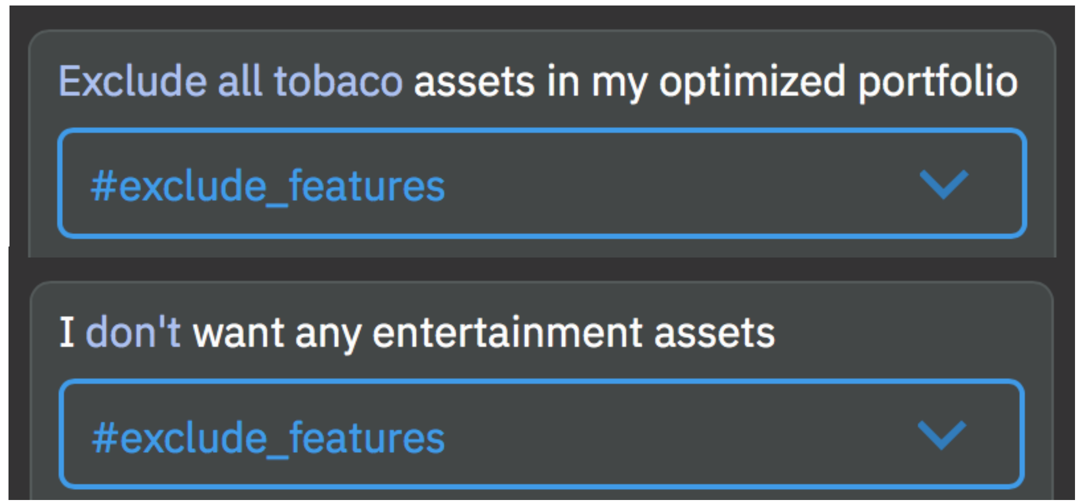


Chat-bot for portfolio selection

Methodology

Watson Assistant
Intents Classifier
(Neural Networks Model)

Watson Assistant: Intents Classifier



These two examples appeared to be different, but they are recognized successfully by the model and tagged as **#exclude_features**

Pre-defined Dictionary

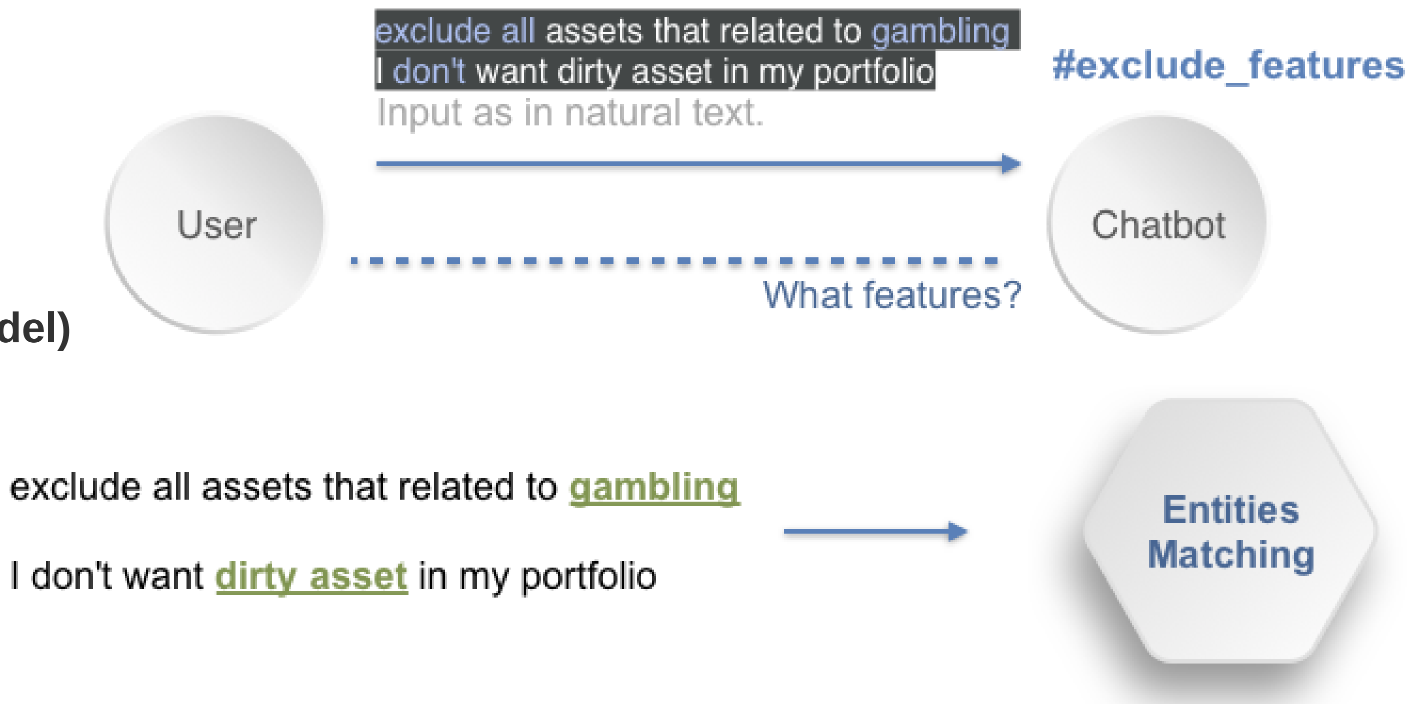
Do not heavily depend on the pre-defined dictionary.

Chat-bot for portfolio selection

Methodology

Watson Assistant: Intents Classifier

Watson Assistant
Intents Classifier
(Neural Networks Model)

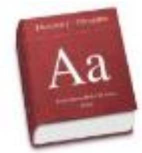


Chat-bot for portfolio selection

Methodology

Watson Assistant
Entities Matching

Watson Assistant: Entities Matching



Pre-defined
Dictionary

Searching for **pre-defined entities** in **precise matching or fuzzy matching**. Fuzzy matching regards changes of part of speech, some common typos as the same (fuzzy). Precise matching generally requires the entities to be appeared exactly as defined.

Currently the Cognitive Portfolio Optimizer Chatbot has 14 types of manually defined entities and **3 system trained entities (number, percentage, currency)**.

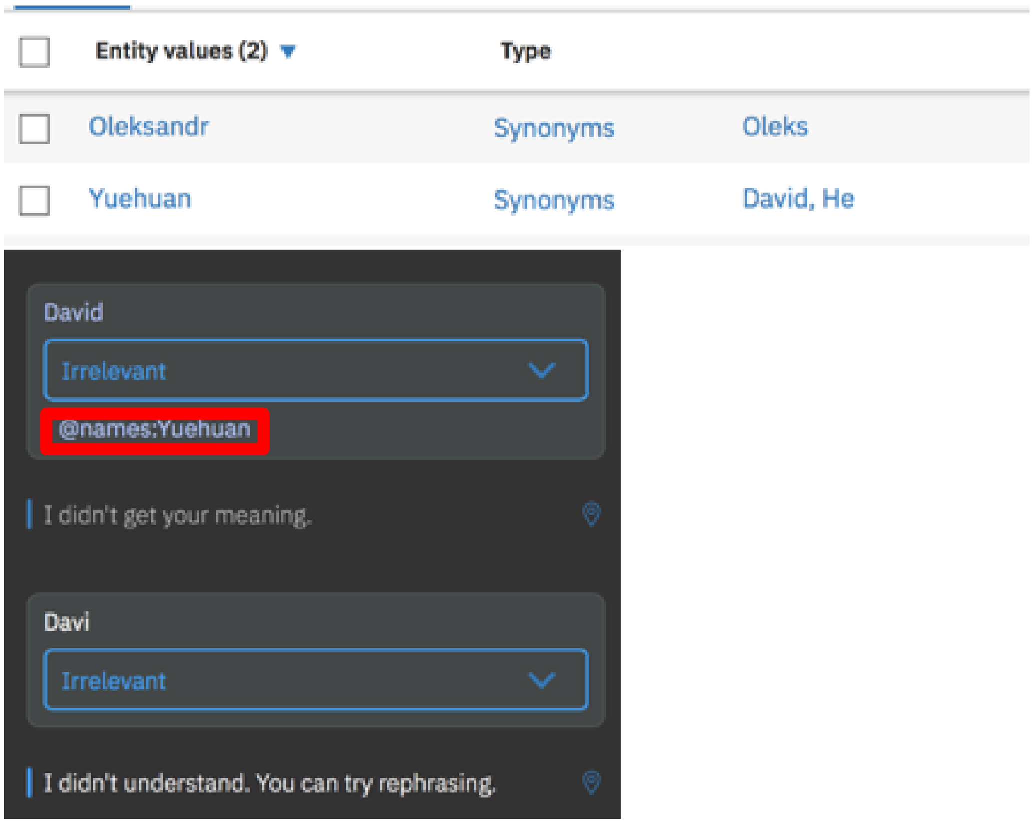
Chat-bot for portfolio selection

Methodology

Watson Assistant
Entities Matching

Watson Assistant: Entities Matching

Precise Matching



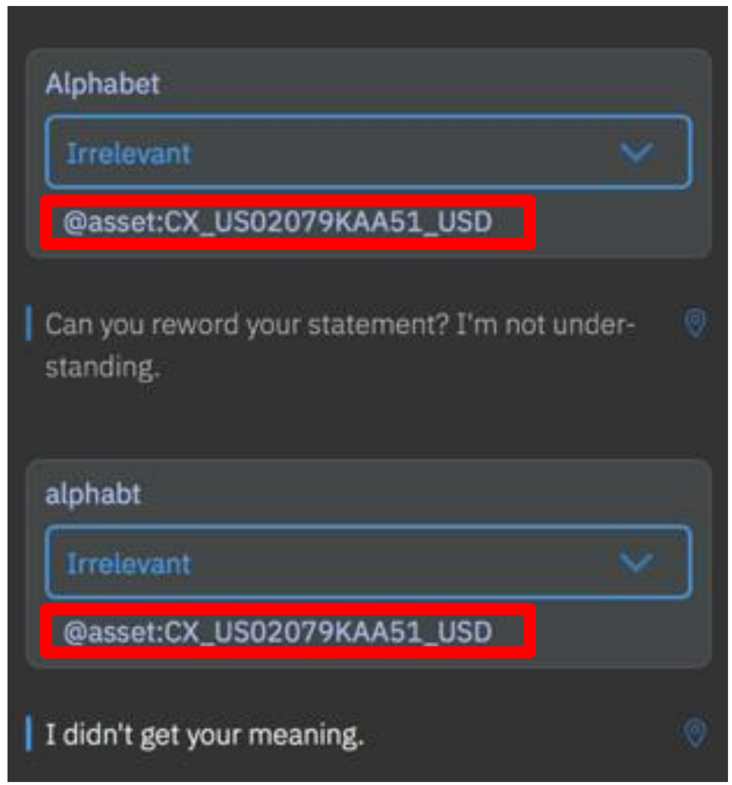
Chat-bot for portfolio selection

Methodology

Watson Assistant
Entities Matching

Watson Assistant: Entities Matching

Fuzzy Matching



Entities type: asset
Entities value:
CX_US02079KAA51_USD

Synonyms:
ALPHABET, US02079KAA51,
02079KAA5,
ALPHABET INC

Fuzzy Matching: enabled

Chat-bot for portfolio selection

Methodology

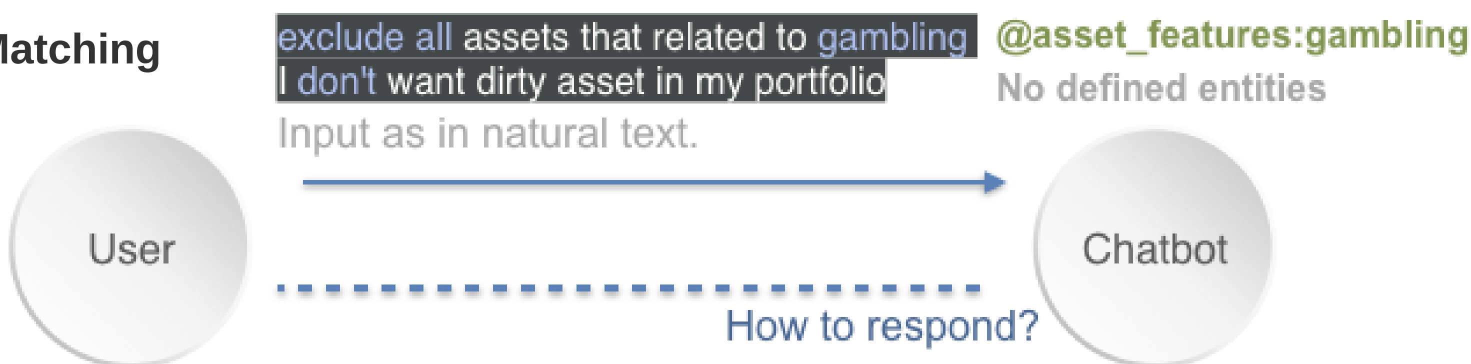
Watson Assistant
Intents Classifier
(Neural Networks Model)

&

Entities Matching

Watson Assistant: Entities Matching

To generate responses, ask for more specific information or give error message, a dialogue flow is employed.

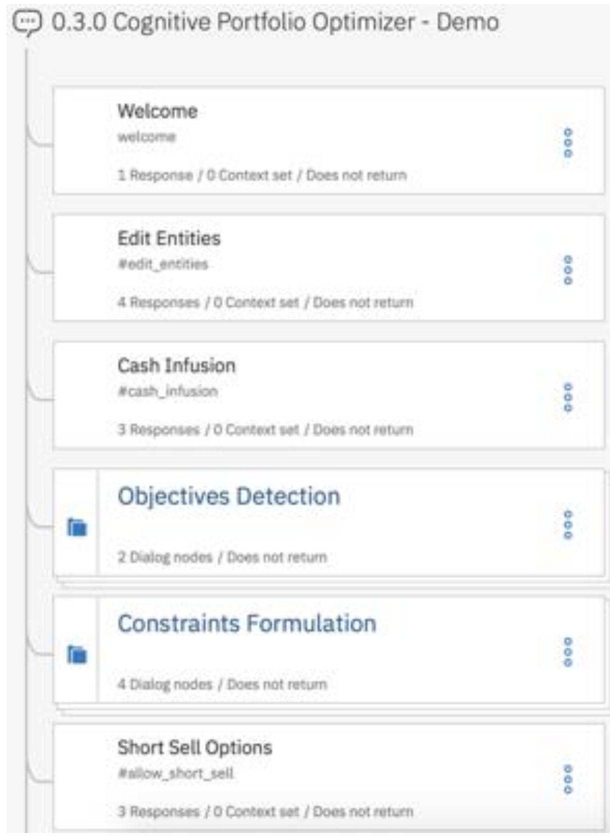


Chat-bot for portfolio selection

Methodology

Watson Assistant

Watson Assistant: Dialogue Flow



Dialogue flow is designed to help chatbot to generate response to the user.

The chatbot will scan each node in the dialogue flow and if its conditions are triggered, the corresponding response will be provided to the user.

Generally, an anything_else node is added at the end of the dialogue flow and each hierarchy level of dialogue flow.

Chat-bot for portfolio selection

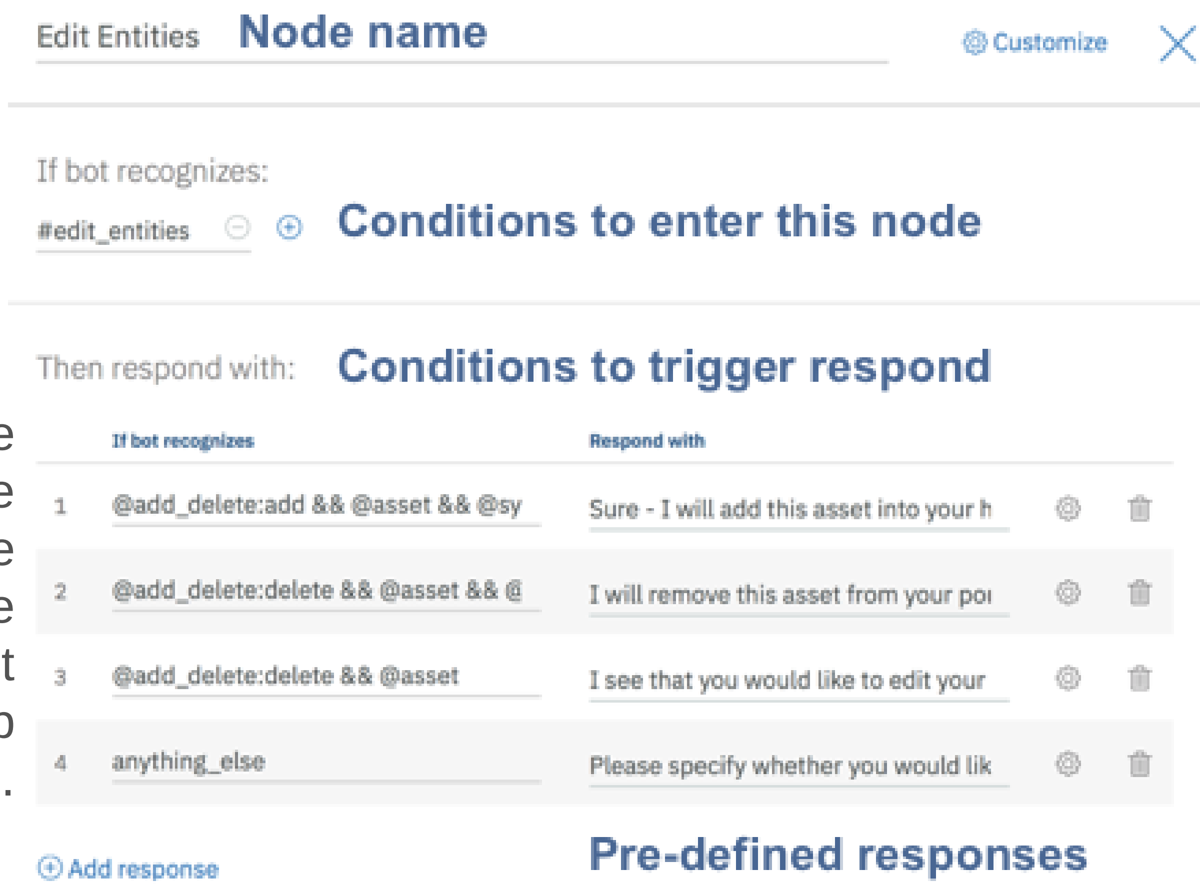
Methodology

Watson Assistant

Sometimes, to help the chatbot appear to be more humanlike, more than one response are provided for the exact same conditions to help varying the responses.

Random / Sequential

Watson Assistant: Dialogue Flow



Chat-bot for portfolio selection

Methodology

Watson Assistant: Sample Dialogue Flow

@add_delete:add && @asset && @sys-number	Sure - I will add this asset into your holdings!
@add_delete:delete && @asset && @sys-number	I will remove this asset from your portfolio.
@add_delete:delete && @asset	I see that you would like to edit your portfolio holdings. Please specify the name and quantity of the asset you would like to edit, along with whether you would like to add or remove this asset.
anything_else	Please specify whether you would like to add or remove an asset from your portfolio, along with the asset name and quantity.

Chat-bot for portfolio selection

Application

Giving Responses and Inquiring Further Information

exclude all assets that related to gambling
I don't want dirty asset in my portfolio

Input as in natural text.

@asset_features:gambling
No defined entities

User

Chatbot

I will exclude all assets with this feature from your portfolio.

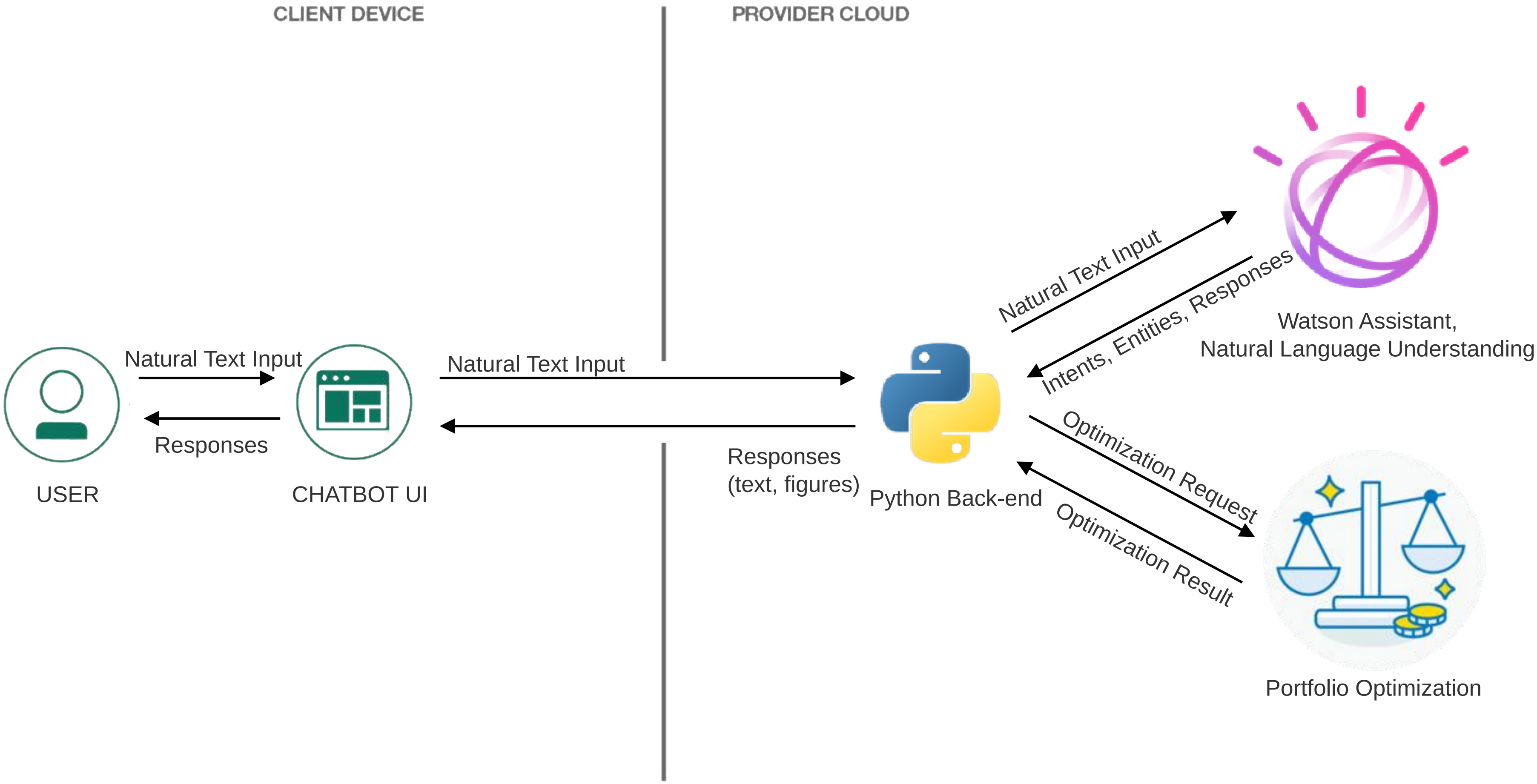
I understand you may want to exclude assets with certain features from your portfolio. Which types of assets would you like to exclude? For example, you may wish to exclude military or tobacco assets.

Python Back-end

Chat-bot for portfolio selection

Application

Cognitive Portfolio Optimization Service Work Flow



Chat-bot for portfolio optimization


```
"objectives": [
 {
   "sense":     "minimize",
   "measure":   "variance",
   "attribute": "return",
   "portfolio": "Universe",
   "TargetPortfolio": "Aggressive",
   "timestep":  30,
   "description": "minimize tracking error squared (variance of the difference between
                   Universe portfolio and Aggressive benchmark returns) at time 30 days"
 }
],
"constraints": [
 {
   "attribute":   "weight",
   "portfolio":   "HighEnvironmental",
   "InPortfolio": "Universe",
   "relation":    "greater-or-equal"
   "constant":    0.5,
   "description": "Creating an average Environmental score of High"
 },
 {
   "attribute": "weight",
   "members":   "Has Tobacco",
   "relation":  "equal",
   "constant":  0.0,
   "description": "Excluding all securities which have the property Has Tobacco"
 }
]
```

Chat-bot for portfolio optimization – demo

☆ **watson-advisor**

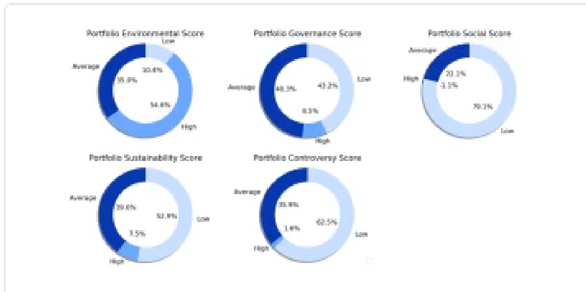
Messages About

Search

Thursday, March 14th

watson-advisor APP 10:46 AM

PortfolioBreakdown



Friday, March 15th

Oleksandr Romanko 1:35 PM

Hi Watson

Today

Oleksandr Romanko 12:32 PM

Hi Watson

watson-advisor APP 12:33 PM

Hi, I'm here to help you build a portfolio of investments with the features you care about. Please tell me your preferences and I'll incorporate them into your portfolio!

Oleksandr Romanko 1:28 PM

lets start from scratch

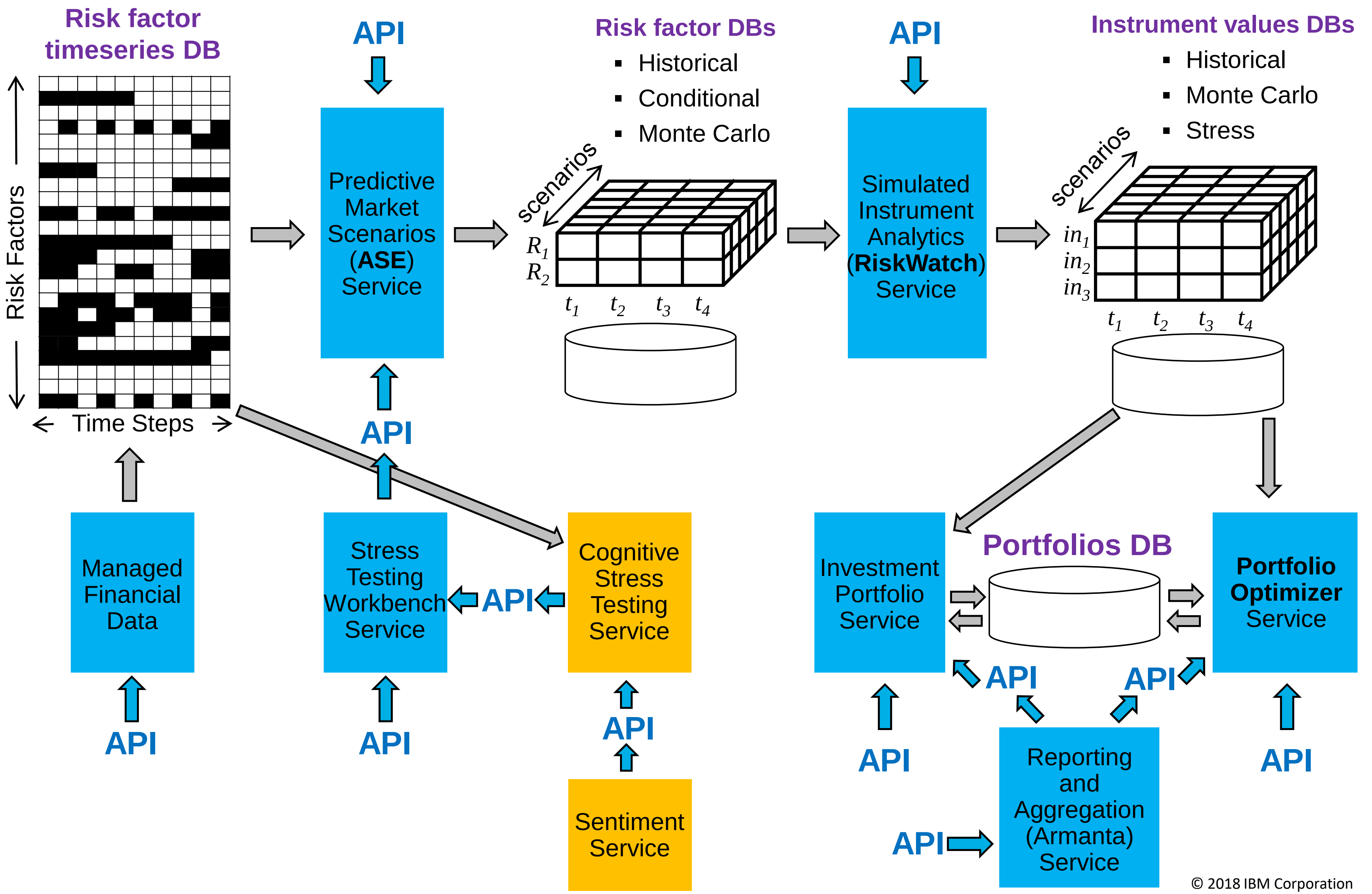
watson-advisor APP 1:28 PM

Sure let's start over.

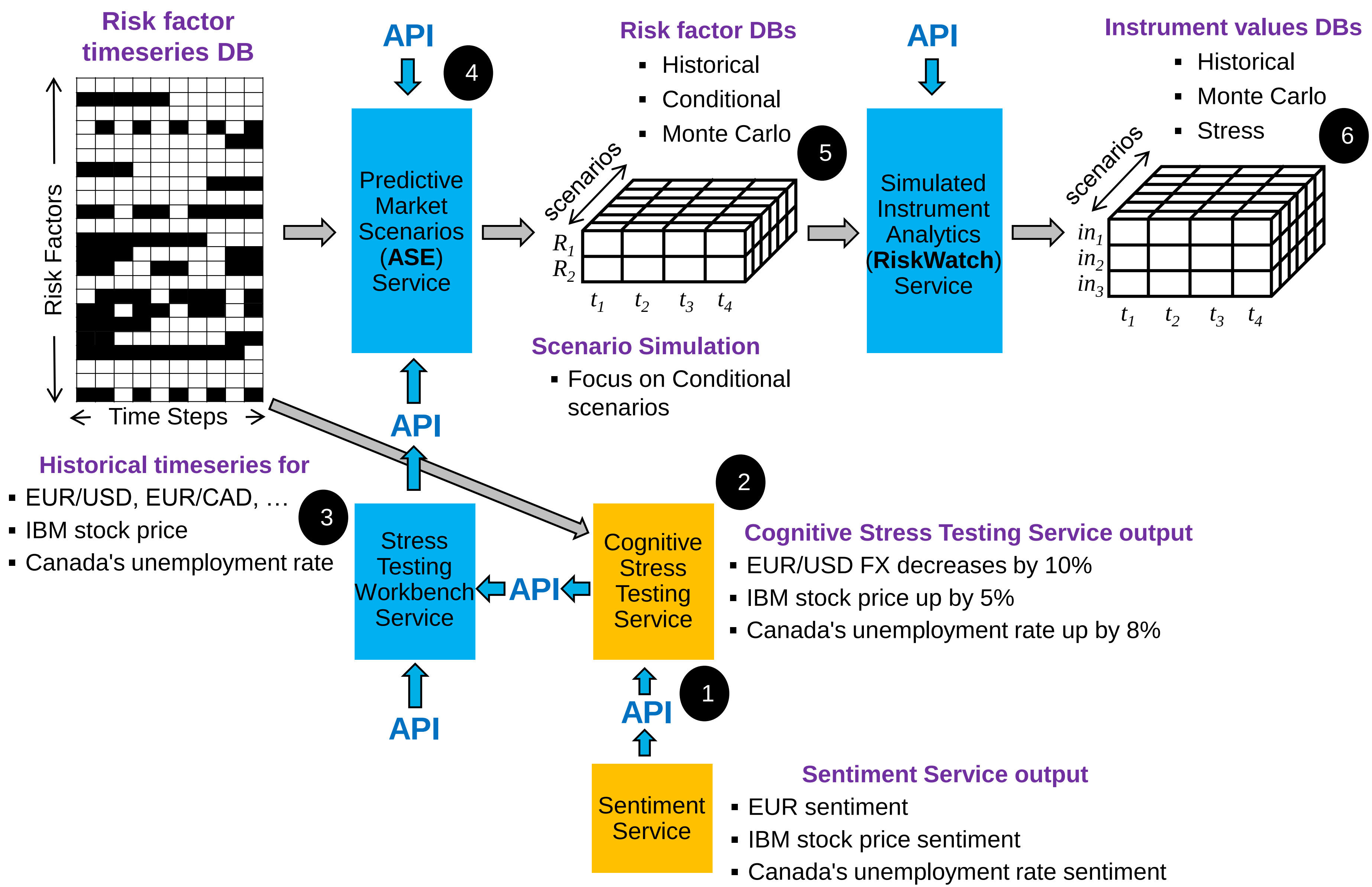
Message watson-advisor

🕮 Notebook 6 – Cognitive Portfolio Stress Testing

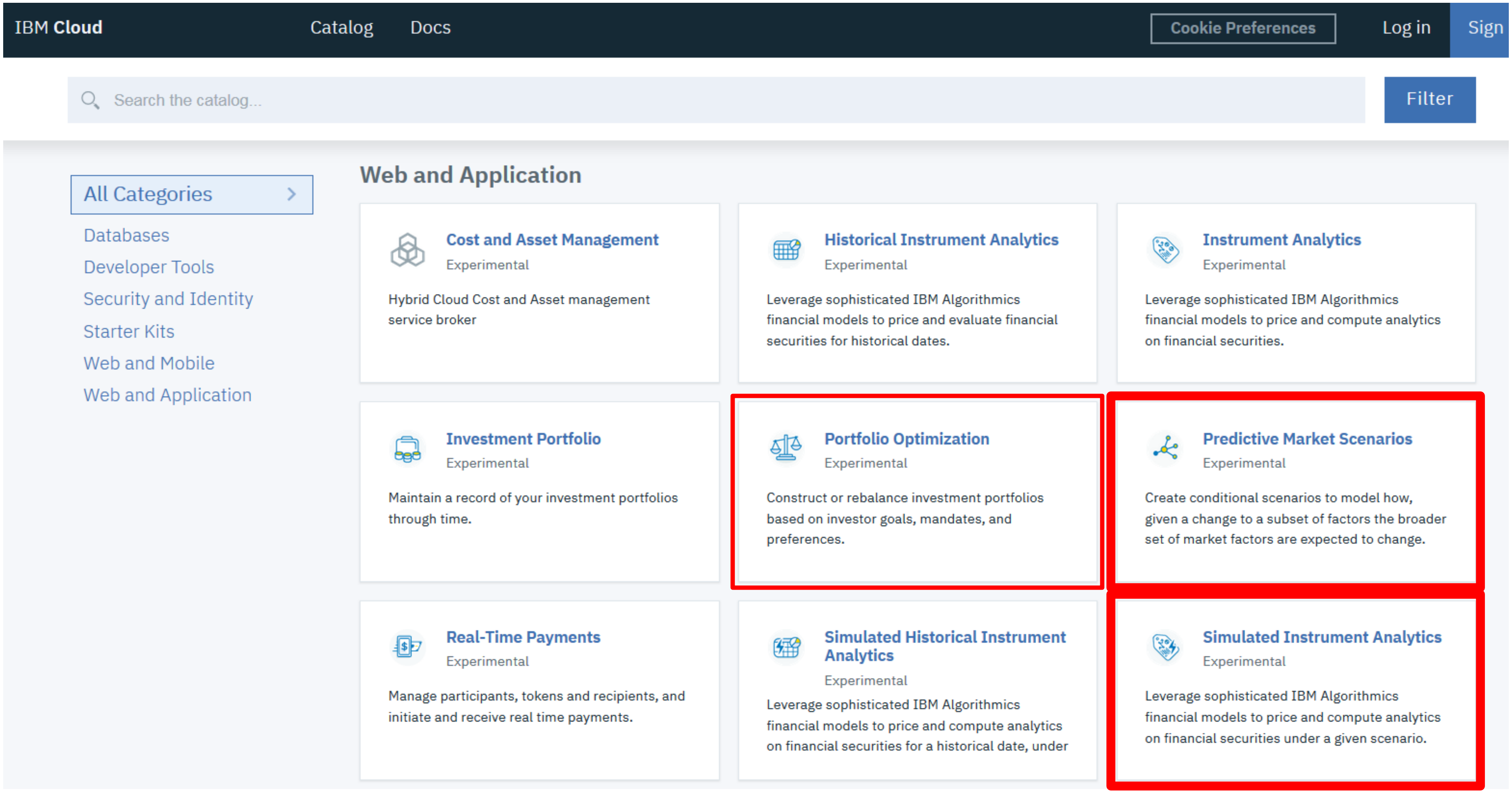
IBM Cloud financial services – cognitive stress testing



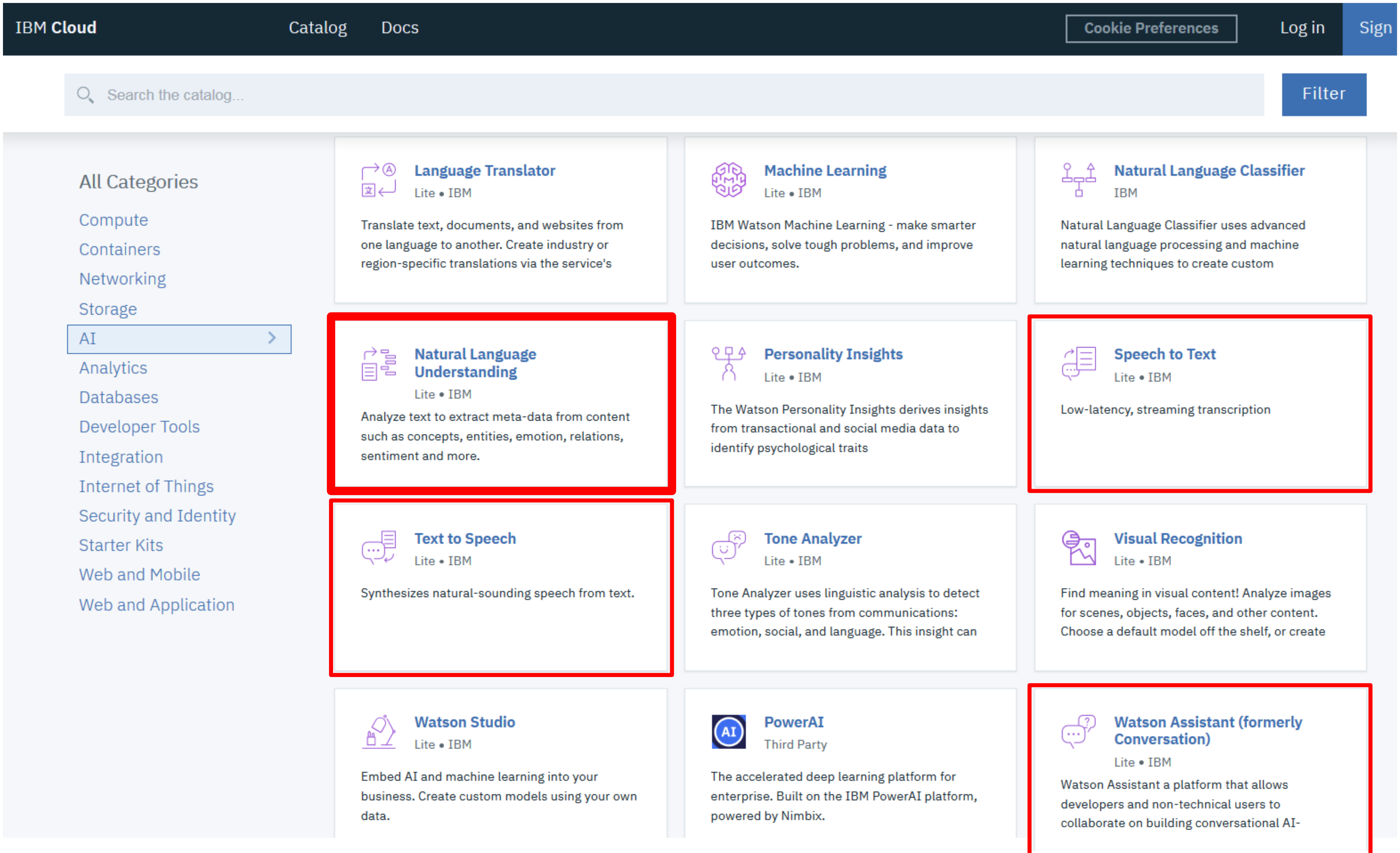
IBM Cloud services – cognitive stress testing



Financial Analytics services on IBM Cloud



Watson Services on IBM Cloud



Risk factor groups in IBM Algo Scenario Engine (ASE)

Risk factor is any observable economic variable whose value, or change in value, may be translated into a change in the value of our portfolio. We tend to categorize risk factors into following risk factor groups:

- Commodities (e.g., Brent oil forward contract price)
- Credit Spread Curves
- Equity (e.g., IBM stock price)
- Foreign Exchange Rate (e.g., EUR/USD exchange rate)
- Implied Volatility
- Interest Rate (e.g., 7-day US interbank interest rate)
- Macroeconomic Factors (e.g., Canada's unemployment rate)
- Market Index (e.g., S&P 500 market index)
- Inflation Rate

Risk factor groups in IBM Algo Scenario Engine (ASE)

Risk factor is any observable economic variable whose value, or change in value, may be translated into a change in the value of our portfolio. We tend to categorize risk factors into following risk factor groups:

- **Commodities** (e.g., Brent oil forward contract price)
- Credit Spread Curves
- **Equity** (e.g., IBM stock price)
- **Foreign Exchange Rate** (e.g., EUR/USD exchange rate)
- Implied Volatility
- Interest Rate (e.g., 7-day US interbank interest rate)
- Macroeconomic Factors (e.g., Canada's unemployment rate)
- Market Index (e.g., S&P 500 market index)
- Inflation Rate

Natural Language Processing: ‘bag of words’ and sentiment analysis

examples (news articles)

	features (word frequencies)					target
	bear	tea	love	bad	drink	sentim
All bears are lovely	1	0	1	0	0	56%
Our tea was bad	0	1	0	1	0	-35%
That bear drinks with bear	2	0	0	0	1	-5%
The bear drinks tea	1	1	0	0	1	4%
We love bears	1	0	1	0	0	63%
	x_1	x_2	x_3	x_4	x_5	y

bag of words

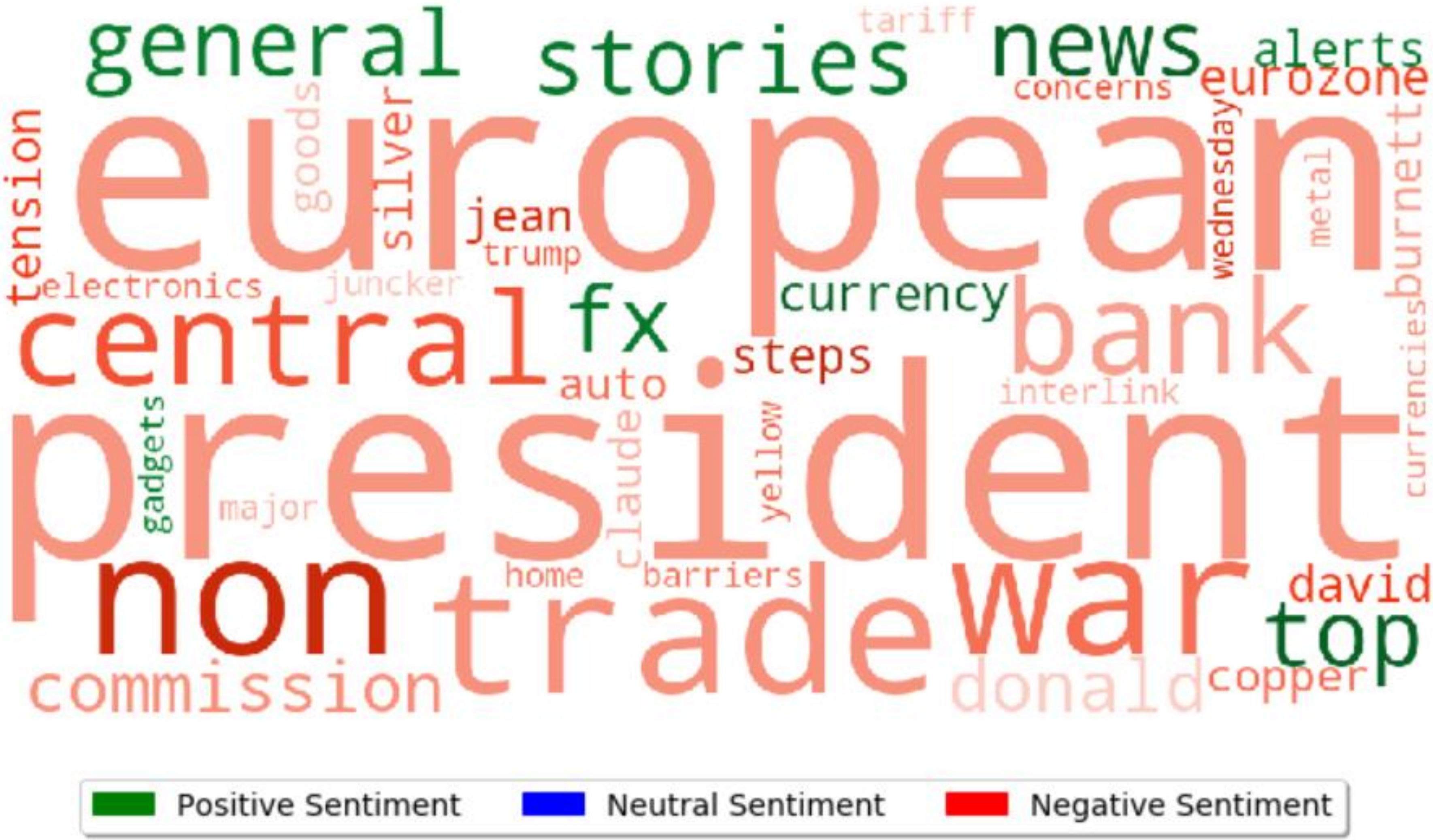
Supervised machine learning algorithm:

- ❑ Linear regression
- ❑ Decision trees
- ❑ SVM regression
- ❑ k-NN regression
- ❑ Ensembles (random forests, XGBoost)
- ❑ Artificial neural nets (deep learning)

$$y = f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f_{\boldsymbol{\theta}}(x_1, \ldots, x_5)$$

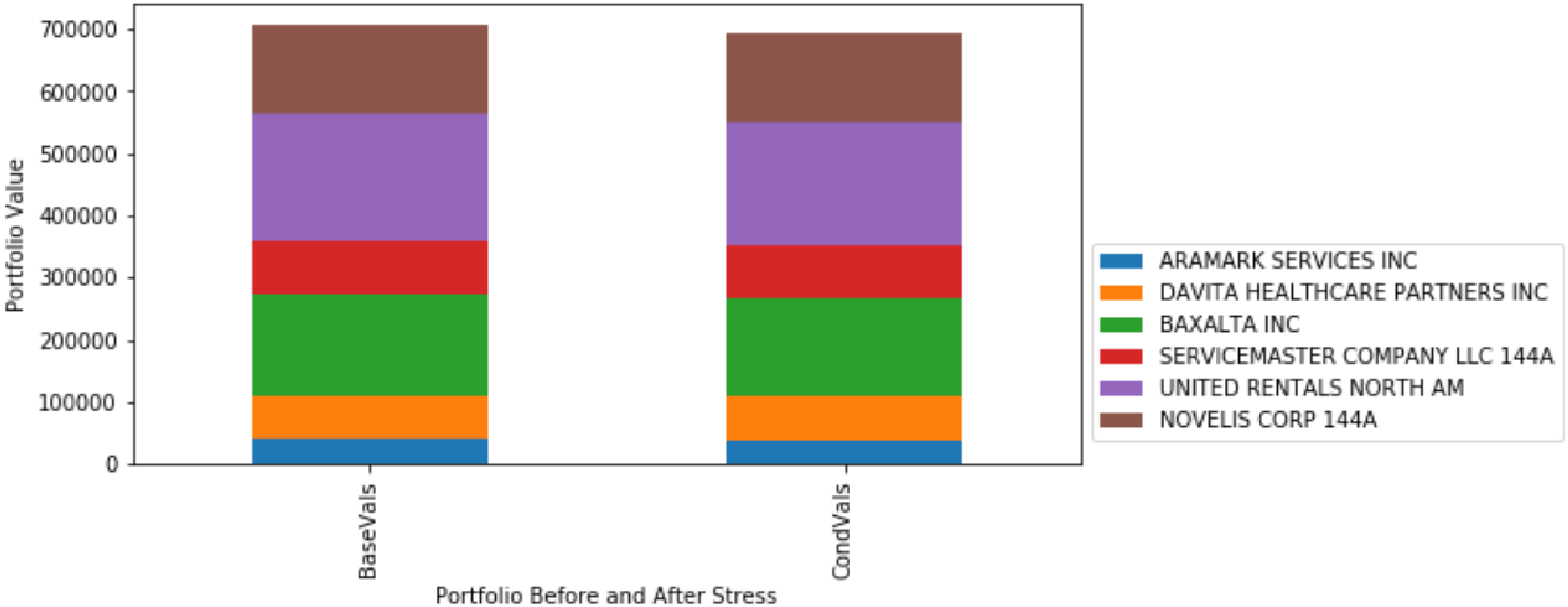
$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \ldots + \theta_5 \cdot x_5 + \epsilon$$

Natural Language Processing: word frequency (Word Cloud)



Cognitive portfolio stress testing demo

```
Portfolio expected value before stress: $707727.48
Portfolio expected value after stress : $692936.21
Portfolio expected value changed by -2.1% under stress
```



Legal Disclaimer

-
- The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.
- References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.
- If the text contains performance statistics or references to benchmarks, insert the following language; otherwise delete:
  Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
- If the text includes any customer examples, please confirm we have prior written approval from such customer and insert the following language; otherwise delete:
  All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.
- Please review text for proper trademark attribution of IBM products. At first use, each product name must be the full name and include appropriate trademark symbols (e.g., IBM Lotus® Sametime® Unyte™). Subsequent references can drop "IBM" but should include the proper branding (e.g., Lotus Sametime Gateway, or WebSphere Application Server). Please refer to http://www.ibm.com/legal/copytrade.shtml for guidance on which trademarks require the ® or ™ symbol. Do not use abbreviations for IBM product names in your presentation. All product names must be used as adjectives rather than nouns. Please list all of the trademarks that you use in your presentation as follows; delete any not included in your presentation. IBM, the IBM logo, Lotus, Lotus Notes, Notes, Domino, Quickr, Sametime, WebSphere, UC2, PartnerWorld and Lotusphere are trademarks of International Business Machines Corporation in the United States, other countries, or both. Unyte is a trademark of WebDialogs, Inc., in the United States, other countries, or both.
- If you reference Adobe® in the text, please mark the first use and include the following; otherwise delete:
  Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- If you reference Java™ in the text, please mark the first use and include the following; otherwise delete:
  Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- If you reference Microsoft® and/or Windows® in the text, please mark the first use and include the following, as applicable; otherwise delete:
  Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.
- If you reference Intel® and/or any of the following Intel products in the text, please mark the first use and include those that you use as follows; otherwise delete:
  Intel, Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
- If you reference UNIX® in the text, please mark the first use and include the following; otherwise delete:
  UNIX is a registered trademark of The Open Group in the United States and other countries.
- If you reference Linux® in your presentation, please mark the first use and include the following; otherwise delete:
  Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.
- If the text/graphics include screenshots, no actual IBM employee names may be used (even your own), if your screenshots include fictitious company names (e.g., Renovations, Zeta Bank, Acme) please update and insert the following; otherwise delete: All references to [insert fictitious company name] refer to a fictitious company and are used for illustration purposes only.