1 About us

Company overview:

- Name: Atomic Intelligence d.o.o.
- Address: Bencekovićeva 33, 10000 Zagreb

Contact person for Challenge Use-Case:

- Name: Tomislav Križan
- e-mail: tkrizan@atmc.ai
- Mobile: +385 98 380 494

Atomic Intelligence d.o.o. is company with headquarters in Zagreb (HR) with primarily focus on EU and USA markets and specialization in areas of data governance, data integration, data warehousing and business intelligence, implementation of data application, data mining, machine learning and artificial intelligence in traditional analytical systems but also in systems using Big Data environments and platforms. Beside services of implementing solution within before said subject areas, Atomic Intelligence d.o.o. (abbreviated AI) is offering services from area of strategic IT consulting, development of custom solutions tailored to specific user needs and expertise in areas of Enterprise and Solution architecture. AI employees have vast and multi-year/decades of practical experience in different industries which can be backed by relevant certificates and customer references.

2 Challenge Use-Case

We are proposing following names for this challenge: **Dynamic Deal Scoring** or **Dynamic Pricing**

3 Challenge Use-Case description

Before innovative pricing systems came along, companies would traditionally look at the prices of similar competitor products and benchmark the value of their own product to come up with the best pricing strategy for their own line of goods. The problem with this manual approach is that although it is manageable for small companies with a reasonable number of products, it is unmanageable for larger companies who stock thousands of items. Each product requires a separate strategy and marketing teams are not able to keep up. it’s hard to overstate the importance of getting pricing right. On average, a 1 percent price increase translates into an 8.7 percent increase in operating profits (assuming no loss of volume, of course). Yet it is estimated that up to 30 percent of the thousands of pricing decisions companies make every year fail to deliver the best price. That’s particularly troubling considering that the flood of data now available provides companies with an opportunity to make significantly better pricing decisions. And that is a lot of lost revenue, which means that potentially millions of dollars are lost in inefficient pricing decisions.

Building advanced analytics such as dynamic deal scoring into the core commercial process helps software sales organizations price smarter, streamline the approval process, and win more deals. Vendors migrating to subscription models, or to software as a service (SaaS,) find that discounts in initial deals are the main determinant of future customer lifetime value.

Traditional software players face increasing discount pressure as they compete with disruptive next-generation players, says the report. Loose discounting practices are hard to rein in:

- Sales representatives are often motivated purely on bookings
- The low marginal cost of software drives an “every dollar is a good dollar” mentality
- The common and myopic quarterly management approach of closing deals at any price, at the end of the quarter, all drive average software discounts to levels not seen in the past

Sales reps often argue that higher discounts are necessary to win deals, but consultancy companies (McKinsey, Deloitte, etc.) research across companies consistently shows the opposite to be true: successful deals almost always have lower average discounts than deals that were lost.

The root of the problem is a lack of insight into objective comparison points. Although sales reps have a good feel for the market, they don’t have much information about how their colleagues’ price similar deals. And to managers, every deal can look unique, forcing them to make approvals based more on gut, or accept the sales reps’ argument that the proposed discount is what it takes to win the deal, instead of an actual objective fact base.

Rewiring the commercial process to leverage advanced analytics for discount management does not only improve commercial productivity and effectiveness—it is also one of the few actions that sales reps actually embrace. Instead of unpopular, top-down corporate discount guidance, reps can use real information to compare their own deals with those their peers are making, which empowers them to make decisions, reduces the red tape they have to deal with, and ultimately allows them to capture a share of the upside through higher commissions.

For pricing and discounting guidance to be effective, it needs to be given at the moment that pricing decisions are made, not just at the end of the process when a deal is being submitted for approval. Typically, this means that deal scoring gets seamlessly integrated into the CRM (customer-relationship-management) or configure-price-quote system.

Goal of this Challenge is to analyze input data for large manufacturing company with locations across the globe (which adds location as one of dimensionality) and look into modeling based on previous sales activities (already won deals and their respective invoices) across multiple dimensions and create model which predicts and recommends best price for every new deal.

Each price should be represented besides numerical value (based on appropriate gross margin percentage over input cost information) with lettering grade A, B, C, D and F similar to schooling grades in US. Lettering grade is used in Dynamic Pricing solutions to better convey message about proposed value of pricing.

Besides price grade with accompanying cut-off (margin level), output result should contain information about attributes which were used to calculate result because in majority of cases, not all input values will be used every time when scoring is done. For example, if model is based on "decision tree" structure, traversing path through that tree should be part of output to enable customer with visual representation of decision process. If model is based on deep learning concepts, abstraction and explainability of model should be done before sending output results.

Because we are talking about solution which would augment existing software at respective client location, deployed Dynamic Deal Scoring model should be exposed as REST API to minimize unnecessary changes over existing software used by Client(s). Based on recommended programming languages in next chapter, this shouldn't be complicated task.

As already mentioned, REST API is here to minimize implementation effort towards end client use-cases (and future extensibility of solutions) and it should support only one endpoint for scoring of incoming data, i.e., we can call it **scoring** where we can send <u>single payload</u> containing our single deal/quote and model in background should evaluate and score data and return <u>single result</u> with all necessary information which will depend on type of model implementation. One example of model implementation is using decision trees where we may return information of which tree node was selected on leaf level as our scoring result and for each node, we should have pre-calculated node cutoffs per each price grade.

In this use-case, we <u>are not focusing</u> on exact end price result but rather on price grade representation (A, B, C, D or F) whit appropriate information about cutoff's (cost vs price uplifts) which tells end user what is expected gross margin in each of those cases.

Quality of solution would be measured on difference between exact gross margin achieved on real invoices vs. predicted by Dynamic Deal Scoring model, and how wide pricing bands are achieved based on appropriate data elements. More narrower bands – i.e., difference between min and max price for each segment – gets better and more precise/tailored results for end user.

For evaluation of delivered models, we will use new dataset with same format of data shared for model training, but which is not previously observer by training process so that we can measure how model behaves when it encounters new product, customer or any combination, so crucial component of each model is finding appropriate balance for our key variables such as deal size, customer product mix, etc.. At same time, impact of each recommended price will be calculated against invoice selected price to show does model tends to uplift or mode focus on price discounts based on different profiles of data, and does it do this in consistent way for single profile.

Note: You do not need to worry about scalability of solution and load balancing, etc., but rather you can focus of quality of price prediction which comes from single execution thread.

3.1 Educative segment of this challenge

Developing solution for this challenge, you will have opportunity to understand and learn how different price selection can interact between themselves in multinational company serving Customer in different continents. You will learn how to prepare data, see if there is need to create calculated/inferred features which are not directly present in provided dataset (like rolling averages of behavior, new calculations from few combined measurements, etc.), discard any unrelated piece of information which can complicate your modeling and output result.

This specific challenge task focuses on area where companies have huge revenue fluctuation in terms of price selection by sales representative who are mostly focused on their short-term goal (like bonus) instead of long-term company earnings. You will have possibility to observe how people are making pricing selections and how that process can be improved by using customer analytics (customer purchase behavior) and product analytics.

There are multiple areas of applying solutions like this which are not tied to specific industry (all are dynamic in nature):

- Pricing solution for Software development company and how to match it to competition – where did they win and where did they lose in price selection
- Pricing solution for web shop retailer – offering best price at time of purchase to maximize retention of customers and minimize abandonment rate on web shop portals
- Pricing solution for B2B segment – large manufacturing companies which are serving other companies can price solution based on seasonality, behavior,
- etc.

4 Recommended programming languages

R & Python as Open Source solutions which can do whole end-to-end processing.

You can find all information for installation of selected programming language online because those depends on your local environment and OS. Some guidelines which can be helpful are:

- R: https://rstudio-education.github.io/hopr/starting.html
- Python: https://realpython.com/installing-python/

We recommend also to read CRISP-DM methodology ((http://www.sv-europe.com/crisp-dm-methodology) to prepare yourself with Data Science approach for gathering information and data, preparing them, modeling, evaluating and deploying.

First two phases of CRISP-DM methodology would be completed by us (**1. Business Understanding** and **2. Data Understanding**) while other steps (except last**: 6. Deployment**) are focus of this challenge.

5 Required Computing Resources

Minimum System Requirements

- Processors: Intel® Core™ i3 processor
- Disk space: 1 GB
- RAM: 4GB
- Operating systems: Windows* 7 or later, macOS, and Linux

Recommended System Requirements:

- Processors: Intel® Core™ i7 processor
- Disk space: 2-3 GB
- RAM: 8-16GB
- Operating systems: Windows* 7 or later, macOS, and Linux

Additional recommendation is to open Cloud trial (AWS, Azure, Oracle Cloud, Google Cloud) and use it for testing and building models if more computation power is required (our case will not require such computation requirements, but there is option to use Cloud environments)

6 Dataset example

Top 10 records for dataset created from previous completed invoices are as follows (example is split in 4 tables to have visible result where every row from one table aligns with same row in other tables – vertical data separation):

Manufacturing Region	Manufacturing Location Code	Intercompany	CustomerID	Customer industry	Customer First Invoice Date	Customer Region	CUSTOMER Geography - State	CUSTOMER Geography - Count	Top Customer Group
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	108879	AEROSPACE	15.9.2005 0:00	North America	CA	UNITED STATES	MEGGITT
North America	N2	No	109037	AEROSPACE	24.10.2008 0:00	North America	NY	UNITED STATES	OTHER
North America	N2	No	109037	AEROSPACE	24.10.2008 0:00	North America	NY	UNITED STATES	OTHER
North America	N2	No	109037	AEROSPACE	24.10.2008 0:00	North America	NY	UNITED STATES	OTHER
North America	N2	No	108879	AEROSPACE	15.9.2005 0:00	North America	CA	UNITED STATES	MEGGITT
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	109313	AEROSPACE	14.6.2013 0:00	North America	ON	UNITED STATES	OTHER
North America	N2	No	108910	DEFENSE	25.1.2006 0:00	North America	MI	UNITED STATES	L3COM

Item Code	Product family	Platform	Product group	Price last modified date in the ERP	Born on date	Make vs Buy
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
20957A	THERMAL	Heat Exchange	AAS Pod HX	17.4.2019 0:00		MANUFACTURED
10106A	THERMAL	Heat Exchange	Engine Oil Coolers			MANUFACTURED
10106A	THERMAL	Heat Exchange	Engine Oil Coolers			MANUFACTURED
10106A	THERMAL	Heat Exchange	Engine Oil Coolers			MANUFACTURED
31305A	THERMAL	Heat Exchange	AHU HX	16.4.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
31368A	THERMAL	Heat Exchange	Cinema Projectors	12.3.2019 0:00		MANUFACTURED
10106A	THERMAL	Heat Exchange	Engine Oil Coolers			MANUFACTURED

Sales Channel - Internal	Sales Channel - External	Sales Channel - Grouping	Invoice Date	Invoice #	Invoice Line #	Order Date	Order #	Order Line #
Unassigned	Unassigned	OEM	4.1.2019 0:00	56131	1	24.8.2018 0:00	15684	1
Unassigned	Unassigned	OEM	15.1.2019 0:00	56197	1	24.8.2018 0:00	15684	1
Unassigned	Unassigned	OEM	15.1.2019 0:00	56197	2	24.8.2018 0:00	15684	1
Unassigned	Unassigned	OEM	29.1.2019 0:00	56285	1	24.8.2018 0:00	15684	1
Unassigned	Unassigned	OEM	29.1.2019 0:00	56286	1	7.11.2018 0:00	15856	1
Sales Rep 2	Sales Rep 2	OEM	13.2.2019 0:00	56390	1	18.1.2018 0:00	15186	1
Sales Rep 1	Sales Rep 1	OEM	22.2.2019 0:00	56498	1	13.11.2018 0:00	15875	1
Sales Rep 1	Sales Rep 1	OEM	22.2.2019 0:00	56498	2	13.11.2018 0:00	15875	1
Sales Rep 1	Sales Rep 1	OEM	22.2.2019 0:00	56498	3	13.11.2018 0:00	15875	1
Sales Rep 2	Sales Rep 2	OEM	7.3.2019 0:00	56615	1	14.3.2018 0:00	15346	2
Unassigned	Unassigned	OEM	18.3.2019 0:00	56678	1	26.10.2018 0:00	15813	1
Unassigned	Unassigned	OEM	18.3.2019 0:00	56678	2	26.10.2018 0:00	15813	1
Sales Rep 1	Sales Rep 1	OEM	30.3.2019 0:00	56786	1	4.10.2018 0:00	15775	1

Invoiced qty (shipped)	Ordered qty	Invoiced pric	Cost of part	Material cost of part	Labor cost of pa	Overhead cost of part	GM%	# of unique product items/groups in a deal	# of unique products on a quote
36	200	544	419,04693	337,59534	22,52488	58,92671	0,229693	1	1
119	200	544	419,04693	337,59534	22,52488	58,92671	0,229693	2	1
1	200	544	419,04693	337,59534	22,52488	58,92671	0,229693	1	1
6	200	544	419,04693	337,59534	22,52488	58,92671	0,229693	1	1
14	39	524,29	419,04693	337,59534	22,52488	58,92671	0,200734	1	1
1	3	28927	18616,86645	13867,56561	1313,38729	3435,91355	0,356419	1	1
2	72	1330	707,03303	295,21883	113,88472	297,92948	0,468396	1	1
6	72	1330	707,03303	295,21883	113,88472	297,92948	0,468396	1	1
20	72	1330	707,03303	295,21883	113,88472	297,92948	0,468396	1	1
1	2	9120	5605,08423	3190,04829	667,86127	1747,17467	0,385407	1	1
3	75	524,29	419,04693	337,59534	22,52488	58,92671	0,200734	2	1
63	75	524,29	419,04693	337,59534	22,52488	58,92671	0,200734	3	1
44	44	1375	707,03303	295,21883	113,88472	297,92948	0,485794	1	1

7 Definition of dataset logical structure

The data will be provided as a structured CSV file.

The columns are:

Name	Data type	Description
Manufacturing Region	String	Global region for manufacturing location (values like North America, Europe, Asia…)
Manufacturing Location Code	String	Manufacturing location unique code
Intercompany	String	Flag which identify does this specific invoice deal with intercompany sale or end customer sale
CustomerID	Integer	Unique identifier for customer
Customer industry	String	Industry to which customer belongs - important for understanding product sold to him and similar groups
Customer Region	String	Global region for customer (values like North America, Europe, Asia…)
Customer First Invoice Date	Date	Date when specific customer had first invoice
CUSTOMER Geography - State	String	Geo component for customer - state information
CUSTOMER Geography - Country	String	Geo component for customer - country information
Top Customer Group	String	Grouping of important customers across locations
Item Code	String	Unique code for item sold
Product family	String	Product family to which specific item belongs
Product group	String	Product group to which specific item belongs. Similar like product family but much fine grained
Platform	String	Name of platform if sold item is part of specific platform (e.g. multiple products sold together as part of whole solution)
Price last modified date in the ERP	Date	Date when price was last changed for specific item sold
Born on date	Date	Date when item was first introduced/manufactured
Make vs Buy	String	Is this part manufactured or bought and then repacked and sold to end customer
Sales Channel - Internal	String	Name of internal Sales representative
Sales Channel - External	String	Name of external Sales representative
Sales Channel - Grouping	String	Sales representative channel
Invoice Date	Date	Date of invoice

Name	Data type	Description
Invoice #	String	Unique invoice number
Invoice Line #	String	Line item number for specific invoice
Order Date	Date	Date of order (order always comes before invoice and it is trigger for invoice when items are shipped/delivered)
Order #	String	Unique order number
Order Line #	String	Line item number for specific order
Invoiced qty (shipped)	Decimal(18,6)	Quantity which was invoiced
Ordered qty	Decimal(18,6)	Quantity which was ordered
Invoiced price	Decimal(18,6)	Unit price we used on invoice
Cost of part	Decimal(18,6)	Unit cost we used on invoice
Material cost of part	Decimal(18,6)	Segment of unit cost - material related cost
Labor cost of part	Decimal(18,6)	Segment of unit cost - labor related cost
Overhead cost of part	Decimal(18,6)	Segment of unit cost - overhead related cost
GM%	Decimal(18,6)	Gross margin percentage - uplift of price over cost
# of unique product items/groups in a deal	Decimal(18,6)	Represents number which are part of a single deal
# of unique products on a quote	Decimal(18,6)	Represents number of items sold together

8 Previous experience over proposed Challenge

We as Atomic Intelligence d.o.o. (company providing Challenge Use-Case) implemented and deployed project which had same or even more complex requirement like described in this Challenge Use-Case.

9 Additional information

To be provided after cases are announced….

10 Mandatory parts of completed solution

Mandatory parts of completed and delivered solution would be:

- Source code for building models, evaluating them and running them – raw code in selected programming language which allows execution and repeatability of delivered solution on different environment (all dependencies should be clearly stated in code)
- Project documentation in form of populated chapters as described within CRISP-DM methodology with focus on what was project understanding, issues with data, premise for modeling (and obstacles if any), selection criteria for training and evaluating models, output results, etc. – please refer to CRISP-DM for full description of process and required steps. – in PDF format
- Technical documentation in form of Run Book which contains information how should delivered solution be run to provide output result as described in Project documentation – in PDF format
- Final presentation in PPT or PDF format

Note:

- All code stays propriety of team (person or persons) who developed it and we as company providing Challenge Use-Case will use it only to validate results of delivered solution (e.g. can we repeat same results as described in delivered solution and user documentation)
- Data used for implementation of project is property of **Atomic Intelligence d.o.o.** and it can only be used for duration of this Challenge.
- Any unauthorized use of data would be strictly illegal and basis for private prosecution under Croatian laws.
- For use of data after Challenge is completed, please contact **Atomic Intelligence d.o.o.** for written permission.

11 Grading criteria

The result should be a clearly presented model that can be realistically implemented. Following criteria will be used to grade solutions:

Criteria	Grade Range	Contributes to Total Grade	Note
Completeness of solution – including project documentation	0-5	35%	Project documentation should contain chapter which cover each of segments in CRISP-DM methodology to describe how team understood requirement, which steps did they considered, discarded and/or implemented for current use-case
Technical documentation	0-5	10%	It should at least have documented all steps of how to run process and get output results either textually or visually
Feasibility of the solution	0-5	10%	
Presentation	0-5	25%	
Quality of the solution	0-5	20%	
	TOTAL:	100%	