CRACKING THE WTP METRIC FOR ENHANCING PRICING FRAMEWORK AND MARGIN OPTIMIZATION USING REASONING LLMS, AGENTS AND LANGCHAIN FRAMEWORK

Chirag Soni [1] and Swati Shah [2]

[1] Senior Product Manager, Data, AI and Insights, PayPal, Bangalore India
[2] Swati Shah, Head of Margin Optimization Product, PayPal, San Jose USA

ABSTRACT

The advancements in LLMs and Agentic Universe has enabled businesses to incorporate intelligent automation across business applications. With this paper we present a methodology to enhance B2B sales process where LangChain framework and Agents could be used to fetch critical information from the internet in real time to indirectly calculate a Business's Willingness To Pay in the form of a score generated by Reasoning LLMs, and rank order these businesses to filter out the ones who at the given point in time have the highest WTP, which could be used by Sales teams to prioritize, negotiate a better pricing and identify sales opportunities currently active in their respective time zones.

KEYWORDS

Generative AI, LLM, LangChain, AI Agents, Reasoning models, Risk, Sales

1. INTRODUCTION

Willingness To Pay is a very complex yet critical metric to assess how and when Sales teams should approach their target companies or businesses in B2B sales process. It indicates how much a Business is willing to pay for the services your company offers, and directly correlates to how you price your products. It is a complex metric because there is no set standardized framework to assess WTP of the Business and the inputs to assess WTP varies individual to individual. Many times the Sales teams would not be assessing the Business on WTP due to its complexity and stand the risk of losing the given deal as the pricing does not work out for both the parties or the deal may be called off due to higher price perception by the target Business. There is also a very critical component in assessing WTP, which is Time. The Business's WTP will be majorly depending on the time during which deal negotiations are happening, and there are multiple inputs that need to be looked at before concluding. Also, even if the Sales representatives have all the required information to calculate a directional WTP, they still would need to gather information from multiple sources which may take several days resulting in delayed sales and onboarding cycles impacting revenue generation. This paper discusses an approach that could be used for B2B sales to assess at a given point in time what would be the WTP for a Business and when is the best time to have the sales process kicked off for these customers. The motivation behind this research was to enable faster Business onboarding and kick start revenue stream sooner since there is considerable time being spent in manual research which is often inconsistent. Also, most pricing decisions consist of data driven strategies to study internal performance and offer products with customized pricing to Businesses based on how well they

are already engaged, and usually does not include estimates of WTP coming from external data and insights. WTP is a metric that's very dynamic in nature and external factors such as Financial health and ongoing market sentiments add a critical element to decipher the WTP. The objective was to automate the entire process to a certain extent by removing dependency on manual research and consolidate all the major factors affecting WTP at one single platform, and uncover businesses which are having highest WTP at given point in time to prioritize sales communications, offering customized pricing to those specific Businesses.

2. METHODOLOGY

2.1. Drivers of WTP

Let's look at the most important inputs that are needed to define and assess WTP. In the next section we will discuss how to consolidate everything and score the Businesses on these parameters to detect WTP –

- Internal to company rank for the Business: for already onboarded Businesses using your products, how has their engagement been on internal to your company's parameters. For e.g. High product penetration, long term relationship, higher profit margins etc. are all metrics that indicate higher rank for these Businesses for within your company performance. For new Businesses that your company is targeting to sell
- Current Financial Health: this indicates how a given Business is performing on Financial metrics like Sales, Operating Expenses, EBITDA, Revenue etc. These are a very strong indicator of how healthy a business is in given point in time
- Market and Investor Sentiments: what are the ongoing sentiments about a given Business in market. Are their customers happy? Are there any news articles indicating customer dissent and risk factors due to the Business's long-term decisions? All these factors are also important to consider when assessing the WTP of a business at given point in time

2.2. Approach – Fetching the Required Data

- Each of the scoring pillars mentioned above will be an input to the final WTP score we will be generating. Below is the framework to prepare for scoring:
- Internal to company Ranking: Performance Score generated over historical transactional data for internal to your company (for already onboarded Businesses) that rates the businesses on all parameters which affect engagement and incoming revenue. The data would seamlessly flow in from internal databases and an algorithm will generate the score at Business account level.
- Financial Health Ranking: this data needs to be fetched from external sources. The most reliable source (for listed companies) is their publicly released Financial statements. The financial metrics discussed in previous paragraphs would usually be available on Google Finance and a trained sales agent would need to fetch this manually at the current time while deal is being framed. They would then need to compare with historical financial data and form an opinion about given Business's health based on individual analysis
- Market Sentiment Ranking: this data should also come from external sources like news articles, press releases, finance outlook articles from data providers like Google Finance etc. A trained sales agent would need to read through this data and understand the performance based on their individual analysis

2.3. Approach – Using LLMs and Agents for Ranking

For the internal to company performance scoring, it would be a direct approach to rank businesses on their performance. However, for pulling external data and analysing the Financial and Sentiment performance is manually intensive activity which can be automated using Agents framework as below

1. Use specially designed prompts with Agents that will call Google Finance API by passing the Business name as input
2. Receive a large Json file with a host of financial metrics for the given Business from Agent
3. Pass this into another pre-created prompt to an LLM. This prompt is designed to summarize financial metrics and consolidate information in a textual format
4. This summary will again be passed into another prompt that will call a reasoning LLM to generate a score based on given information and rank the Business on a predetermined scale
5. Similar call is made to News APIs, and the articles are summarized to build Market Sentiment score by repeating steps 1 through 5

2.4. Approach – Combining Everything Together

Now for the final steps, we will need to consolidate all scores generated from above steps. This is not a direct average as there will be different weightage assigned to each scoring pillar. Based on our research, a Business's current Financial Health compared with historical metrics is one of the most important parameters to assess current WTP as good financial health is direct indicator of ability to pay higher, hence this gets the highest weightage. Next will be the internal to company performance score since even though Business is in good financial health, it may be multi-homing onto different platforms, meaning product penetration may not be with a single company. Therefore, it is important to compare Financial Score with Internal Score and a large gap between both may indicate one of the two things below:

1. High Financial Health and Low Internal to Company Performance means Business is multi-homing and a better deal with your company would be beneficial. It also means that we might be leaving money on table while negotiating with the Business.
2. High Internal to Company Performance Score and Low Financial Health Score indicates high Risk. Even though Business is loyal customer and prefers your company over others, poor financial health is a leading indicator of upcoming pricing challenges where the Business would want more discounts and would be cutting costs from wherever they can, including contracts with your company.

Now what remains is the Market Sentiment Score. Why this is important? Because this is an indicator of futuristic engagement as financially healthy businesses with negative investor sentiments directly contribute to challenges with relationship in future. Although it is not as important as the Internal and Financial score, this is definitely worth considering to gauge WTP more accurately.

Based on above context, we have come up with an estimate of the equation that would work best for calculating WTP of a business at any given point in time:

*WTP = (0.65 * Financial Health Score (1 to 10)) + (0.25 * Internal to Company Performance Score (1 to 10)) + (0.1 * Market Sentiment Score (1 to 10))*

This equation works as a best case estimate that incorporates 360-degree overview of the Business's health, both internally and externally.

3. ESTABLISHING THE PROCESS

The above equation will be used to rank order your customers on a 30-point scale. It does not indicate any amount or percentage but is just an indicative score. But how do we use this? And for what use cases? Below points provide a detailed context of how and why we should use this equation in the pricing process:

3.1. Reading the Score

WTP score acts as an AI based assistant to prioritize pricing conversations with Businesses. Without this score, there will be inconsistent and incomplete pricing frameworks where mostly the internal to company data will be used to drive sales whereas extremely critical external metrics will be completely missed. This also helps rank order onboarded Businesses to identify most important ones at a given point in time

3.2. Generating on an Ongoing Basis

- This score needs to be kept updated constantly. However Financial Statements are released only on a quarterly basis. So in order to be consistent, this algorithm can be run on a quarterly basis across all three scoring pillars
- LLM based summary from Financial and Market sentiment prompts is also a required data point since it will help in understanding why the WTP is high or low
- Due to cost implications from the LLMs, sales representatives should not be allowed direct access to search a Business name and run algorithm in real time to generate the WTP score. It should be auto generated each quarter and stored in a common database within the company

4. BUSINESS USE CASES

This Score will have multiple use cases within the companies as follows:

1. Critical metric for the Risk teams to keep a check on customers and identify leading indicators of Margin challenges – any significant score drops in Financial and Sentiment health of businesses may indicate upcoming challenges in Business' performance, thereby helping them to take actions and mitigate risks in advance
2. Sales teams can use this to prioritize customer conversations – from the list of Companies that you do business with, which are the ones that have a good WTP as of today, so that conversations and sales processes could be prioritized and targeted for them
3. Pricing teams can determine which price points for their products work best, for e.g. If many customers are scored low on WTP then it could indicate high price perception
4. Leadership teams to assess which Markets are healthy by plotting the scores in regional graph and identify new opportunities to grow the business

These are a few direct use cases, however there may be more based on usage by teams within the company.

5. SCORE GENERATION

5.1. Internal to Company Performance Score

Ranking businesses based on internal to company performance will majorly involve three types of data sources:

1. Transactions history
2. Call Center requests
3. Product Penetration

Above datasets need to be combined to generate a performance score using below framework:

- Transactions Data can be used to generate RFM score (Recency Frequency Monetary). This includes looking at how recent the transactions from this business have happened, what is the frequency of transactions coming in, and what has been the monetary value involved.
- Customers reaching our for Service requests indicates their active engagement. This is also a direct indicator of internal to company performance.
- How many products that you offer have been purchased or integrated by the Business, or the increase in penetration indicates growth and good internal performance

5.2. Financial Health Score

This involves multiple steps which could be automated via LLMs, Agents and LangChain framework as below

LangChain and Agents Framework
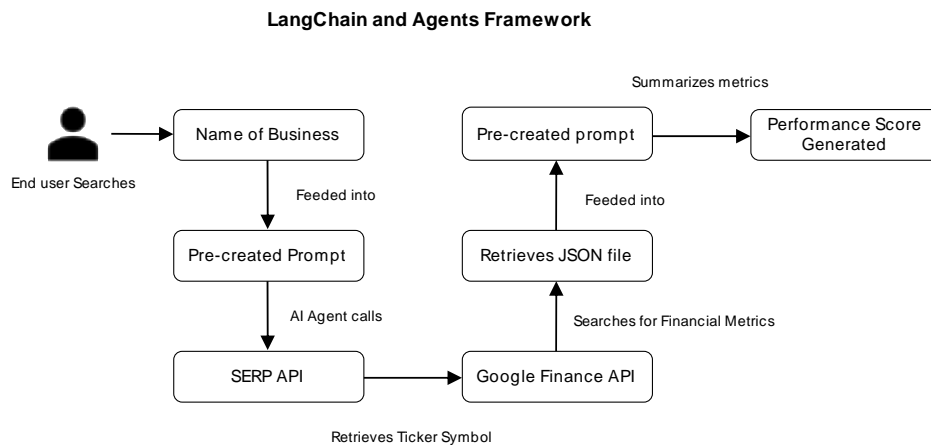
The pre-created prompts form the LangChain where one output from a prompt is feeded as an input to another prompt. Ranking the Business will be done using Reasoning LLMs like GPT o1 which are capable of using Chain of Thoughts on the financial metric summary provided via the previous prompt and then generate a final score.

The score will be updated once a quarter based on financial statement releases

5.3. Market Sentiment Score

The framework is same as above, where instead of calling SERP API for company Ticker symbol and then Google Finance API for financial metrics, the agents will call API from news service providers. These news articles will then be summarized to pass onto the reasoning LLM which will then generate the final Sentiment score.

5.4. Consolidation and Data Repository Creation

Final step will be to feed all scores and summaries into backend tables from where users will be able to access the Scores and use them to track performance. Internal teams can be provided access to this data for their use cases.

6. COST IMPLICATIONS

- Major costs will be incurred from the usage of LLMs where prompts for agent calls will be adding up to the costs, and the retrieved data from Google Finance will be huge.
- Secondly there will be charges from API providers (Google Finance, News Vendors) per call that needs to be considered
- Lastly the infrastructure costs for storing and maintaining the consolidated database will need to be added.

7. TRADITIONAL APPROACHES COMPARISON

Below are some of the traditional approaches for WTP assessment (reference https://postandparcel.info/118470/features/parcel-features/predicting-willingness-to-pay/):

7.1. Manual Survey Based Market Research

Sales teams reach out to their prospects with a questionnaire that provides a rough estimate about the maximum price any given customer will pay for your product (the technique is called Van Westendorp Pricing Model and assessment is performed using 4 questions as described in this article https://sawtoothsoftware.com/resources/blog/posts/van-westendorp-pricing-sensitivity-meter) and this information is then used to set up a price point for the product. This provides only an estimate and is dependent on the sample taken for survey, which may include inherent bias.

7.2. Data Based Approaches

Transaction Price Analysis and Customer Choice Analysis are other techniques which make use of data to provide a rough estimation of customer's WTP by uncovering the specific characteristics of customers that lead to accepting a specific price point for the product in consideration.

Both methodologies discussed above are driven by individual choices and not from the perspective of the business as a whole. Since we are discussing about WTP of B2B customers, these techniques would have limited scope as these would not take into consideration the ability of Business to fund new sales or invest in new products for business growth based on their Financials and P&L growth over time.

The methodology we propose will enable Sales teams to help prioritize which Businesses are currently having that capability at that given point in time (using the scores to rank order Businesses) and provide a fair estimate of expected future growth that would be critical in driving sales conversations. Reasoning LLMs have the capability to apply chain of thought to a given command and retrieve results as a human would do, making the scoring closer to being accurate reflection of manual analysis of the data by a human being, but since this needs to be done for thousands of businesses to rank order them on the score and repeat this exercise on a continuous basis to keep the scores refreshed, it is beyond human capability.

8. LIMITATIONS OF LLM BASED APPROACH

8.1. The most important consideration we had discussed earlier is that this approach works only for publicly listed businesses where financial data is available to consume via any API service provider. Nonregistered business' financial data availability is a challenge and there are no reliable sources for trustworthy information.

8.2. Market Sentiment Scoring is a highly complex process where a lot of training and maintenance of models will be required. There are possibilities of bias in inferring sentiments from news articles, but we will need to train the models to look for signals that can cause financial distress to the Businesses, for e.g. High refund rates from customers, extremely poor product reviews, filing for insolvency proceedings, high tax implications etc.

8.3. The entire project is based on calling external APIs and inferring the information via LLMs. Because of this dependency, the scores would not be generated if API calling fails or there are data outages from the data providers, or in case there are new restrictions put in place while dealing with specific vendors

8.4. As a known fact, LLMs are subject to hallucinations. Although the methodology does not generate new data but tries to infer from existing data provided to the LLMs, still there may be cases where a LLM could incorrectly infer the financial growth of a Business by focusing more on certain parameters and ignoring others. There are techniques to train the model against this bias, but these are not 100% efficient.

9. OVERVIEW WITH EXAMPLE

Let's look at an example output from the model for latest quarter's data. There are 4 columns, first 3 for the three pillars of scoring we have discussed about and the last one for the Overall Score based on equation we discussed. The last two columns are an output from the LLM based scoring algorithm based on how these companies have performed (scores are directional and have been generated based on their financial and news data from Google Finance using LLM prompts). However, the first column is dummy data and depends on how well engaged these two businesses are with your company (let's suppose this is generated by the Internal Performance scoring methodology we discussed), let's say if Microsoft has a very good product penetration and engagement with your company and scores good on other parameters for internal scoring, its Internal Performance Score will be high (row 2) affecting the Overall Score for WTP:

	Internal Performance Score (depends on Company to Company)	Financial Performance Score Q4 2024	Market Sentiment Score Q4 2024	Overall Score
Microsoft	5	8	7	7.15
Microsoft	9	8	7	8.15
Intel	5	4	5	4.35
Intel	9	4	5	5.35

Now suppose Microsoft has not had a good engagement with your company that quarter (row 1), then the overall score will reflect that and score Microsoft lower on WTP, meaning even though Microsoft has a good health as a company, they would have less willingness to pay for your products given they had a decrease in engagement with your company in that Quarter.

Similarly, we see Intel has a low score on both Financial and Market Sentiment health given the challenges company is facing, and even if they have a good engagement with your company based on Internal Performance scoring, it will always be ranked lower than Microsoft since it is penalized for inconsistent market performance.

The same score for Businesses could be generated over continuous time (four quarters for e.g.) and we can assess the overall performance by comparing couple of data points together. This score exposes other aspects of the business that are critical to be looked at while negotiating a deal with them, rather than just relying only on the Internal Performance.

These scores could also be plotted across regional graphs to detect the hotspots for Sales and get leading insights into which regions could pose a challenge in near future if many Merchants seem unhealthy on external parameters.

10. COST EFFECTIVENESS COMPARED TO MANUAL RESEARCH

There are several inputs to Manual Research for getting a directional WTP. Browsing through financial statements, news articles, building a database of metrics etc. involves time, and this is one of the most critical cost components in manual research – Time to Onboard, which directly impacts Revenue Generation. The research may take several days and the Business is waiting to be onboarded to your product till then. There is also opportunity cost of Sales Representative Bandwidth that goes into research which could have been used for ongoing deal conversation with the Business otherwise. Deploying additional resources is costly as it involves training.

The methodology discussed above takes away challenges in Time to Onboard as it reduces the cycle time for research and enables faster Revenue Generation from the Business for your company. This in turn reduces the dependency on Sales bandwidth and covers the costs for additional resources deployed for same research activity across Businesses. With LLMs getting advanced by the day, they are also more cost effective and data scientists can choose from a variety of models that could solve the purpose at lesser price than traditional LLMs.

Based on our extensive experimentation running the methodology discussed, it costs ~ $ 0.02 per query (LLM used – GPT 3.5 Turbo), which includes the prompt, incoming Json file from Google finance, tokens generated for summary and score generation prompt. This would mean ~ $20 for 1,000 Businesses searched and scores generated. Repeating this exercise for the same businesses for 4 quarters will require $80 as costs from the LLM. This does not include costs from API providers or infrastructural costs to store and maintain the data, which depends on a lot of factors.

11. SCOPE FOR IMPROVEMENT

11.1. This methodology is limited to publicly listed Businesses, however that forms a small component of prospective leads for Sales teams in B2B framework. A big area of improving the scope is to include non-listed companies by identifying reliable sources of Financial metrics information about them

11.2. Social Media feedback and sentiments about the Business (directly from its end Customers) are not included in the methodology. This could be another critical input to the WTP, as sometimes leading indicators of financial trouble with a Business often come from social media reviews

11.3. Improvements in scoring mechanism – adding more reasoning elements to the scoring framework could be a good improvement, for e.g. looking at more historical data and addressing the gaps in Financial and Sentiment score using market conditions prevalent during those time periods can help an advanced reasoning LLM to decide whether to penalize the Business for adverse financial performance or to remain neutral since the market conditions themselves are non-favourable for that Business.

12. CONCLUSION

WTP is a complex but critical metric which is often ignored due to the manually intensive processes and frameworks. This methodology enables Sales teams to quickly assess the WTP metric and make conclusive and efficient decisions. The costs are justified given the enhancement to overall process this would bring in and provide ability to look at multiple angles since we can track the changes in WTP at both micro and macro level.

The methodology proposed above automates accurate detection of WTP for any Business at a given point in time by looking at three major pillars of information – Internal to company performance, Financial Health performance and Market Sentiment performance. These three parameters are consolidated together to generate a single score that reflects true assessment of WTP for that business. The process involves automation of external insights (finance and sentiment) integration to internal performance and does this through a weighted average equation discussed through the paper. This score could be used across multiple process to enhance existing logic of Business assessment for customized offerings and also help prioritize Sales teams in driving conversations with the Businesses.

AUTHORS

Chirag Soni is an AI product Manager in Data and Insights team at PayPal and has worked as hands on Data Scientist in previous roles

Swati Shah is the head of Margin Optimization product team at PayPal and has varied experience across digital payments and products, and leading AI teams