

Unlocking Efficiency through Data-Driven Support Insights

Using text analytics to improve customer support quality and team efficiency

Business need

Support team time increasingly spent on repetitive tickets

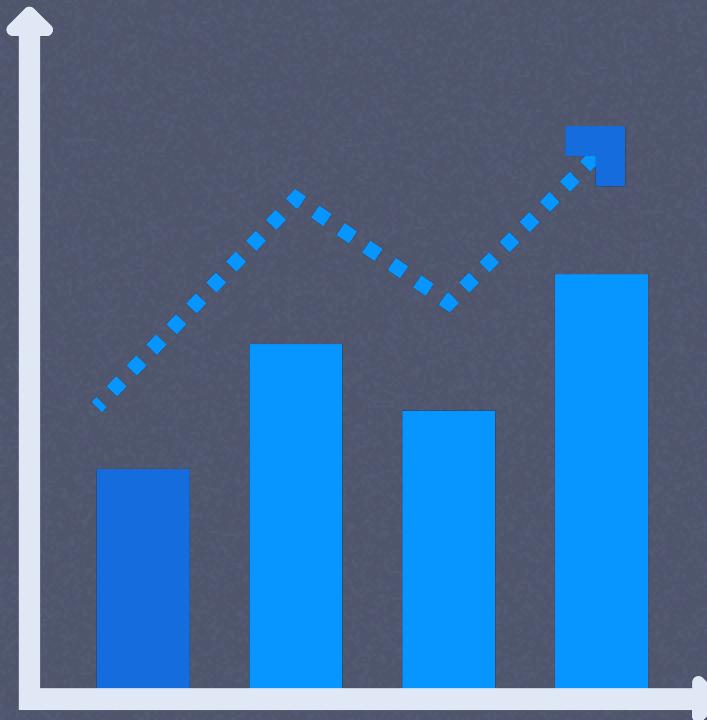
Data Science approach

Cluster analysis of 5 000 chat logs

Expected impact

Reduced workload, faster responses, improved retention

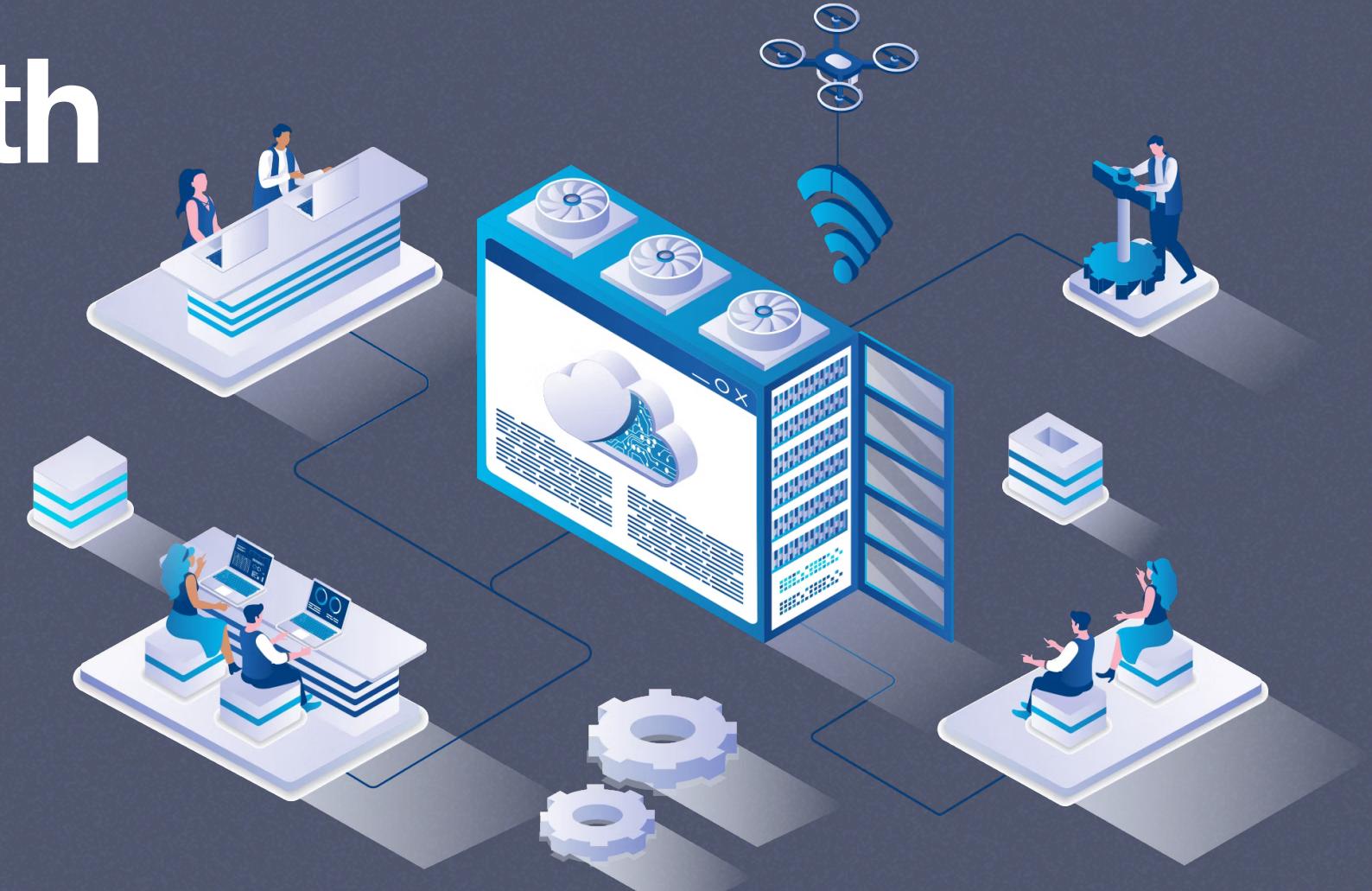
Why Support Efficiency Matters for Growth



Business Context

Project made for European VPS hosting company recognized for its premium, engineer-led customer support.

The support team is composed of highly skilled system administrators who handle both simple and complex technical requests directly, maintaining the company's reputation for quality and trust.



Current State & Baseline Estimates

As this is a boutique hosting provider with a small but highly technical support team, detailed analytics have not yet been formalized.

However, a preliminary internal review and engineer feedback suggest the following patterns for the past 6–12 months:

Metric	Estimate	Source / Note
Tickets per month	150 – 300	Internal ticket logs
Repetitive / low-complexity tickets	25 – 35 %	Engineer feedback + sample log review
Median first-response time (FRT)	20 – 25 min	Internal average (non-urgent queue)
Engineer time on repetitive issues	≈ 25 %	Industry average (Zendesk / Intercom benchmarks)

Source: preliminary internal review + industry averages/benchmarks; exact figures will be refined via exploratory data analysis in the upcoming clusterization stage.

Current Challenge

Hosting's client base continues to grow, the number of incoming tickets has increased significantly.

Engineers now spend a substantial portion of their time on repetitive, low-complexity issues - SSL renewals, DNS records, or migration requests - leaving less availability for critical incidents and system monitoring.

Problem Statement

VPS hosting needs a **data-driven approach** to analyze and categorize recurring support requests, helping to:

- identify the most frequent and time-consuming issues,
- update FAQs and client resources,
- optimize internal workflows to maintain quality while improving efficiency,
- provide groundwork for future intent-based ticket routing or AI-assistant integration.

Why Data Science Matters for VPS Hosting

Hosting's ticket data already contain valuable, untapped insights.

Patterns in wording, timing, and topic recurrence can reveal hidden inefficiencies that humans sense but cannot quantify at scale.



By applying *text analytics* and *unsupervised learning*, the company can cluster similar requests, detect emerging technical issues early, and prioritize documentation updates that reduce repetitive tickets.

Across the hosting industry, data-driven support systems—used by providers like DigitalOcean and Hetzner—have proven effective in lowering response time and improving customer satisfaction. For our hosting, adopting a measured, insight-first approach aligns with its strategic goal:
maintaining premium, human support while using data to enhance efficiency.

Proposed Data Science Approach

We apply text analytics and unsupervised clustering to uncover recurring themes in support tickets.

The goal is to transform historical support logs into structured, interpretable knowledge for product and support teams.

Pipeline

- 1 Preprocessing: language detection, stopword removal, lemmatization, anonymization (e.g., emails, IPs).
- 2 Feature representation: TF-IDF vectors combined with sentiment scores and named entities (e.g., countries, technologies, error types).
- 3 Unsupervised learning: clustering with K-Means, HDBSCAN, or LDA to identify recurrent topics.
- 4 Post-analysis: assign topic labels, extract key terms per cluster, and visualize temporal patterns (issue frequency over time).
- 5 Validation: assess cluster coherence with silhouette and topic-coherence scores and have Support Lead validate top clusters.

Outcome: Transform unstructured text into **actionable insights** — clearer FAQs, smarter routing, and data-driven improvement of support quality.

Future step: extend the model with sentiment and intent detection to enable automated ticket triage.

Alternative Approaches Considered:

Manual tagging or rule-based classification could identify frequent requests but are limited in scalability and adaptability.

By contrast, an unsupervised text analytics approach allows us to surface evolving patterns automatically, enabling continuous improvement without extensive manual effort.

Business Impact & KPI Projection

To quantify the potential value of hosting's data-driven support initiative, we compare realistic pilot goals with industry benchmarks from customer support analytics reports (Zendesk, Intercom, HBR, Freshdesk).

Metric	Industry Benchmark	Hosting Pilot Target	Measurement Approach
Ticket Deflection	Well-structured knowledge bases reduce ticket volume by 25–40 % in 6 months (Zendesk, Intercom)	-15 – 25 % repetitive tickets in covered clusters	Compare ticket tags before / after FAQ updates
First Response Time (FRT)	AI-assisted or optimized workflows lower response time by ~20 % (HBR 2020)	-10 – 15 % median response time in pilot topics	Measure median FRT for clusters before / after
Self-Service Quality	“Helpful article” rate increases by 20 – 30 p.p. after FAQ redesign (Zendesk)	+20 – 30 p.p. on new articles from clustering insights	Track net votes / views / search success
Engineer Time Saved (internal metric)	–	≈ 15 % reduction of time spent on repetitive issues	Formula - hours_saved = deflected_tickets × AHT / 60
Customer Satisfaction (CSAT)	Faster replies boost CSAT by 4–6 points (HBR)	+1 – 3 p.p. CSAT on addressed topics	Compare CSAT scores before / after implementation

Sources: Zendesk 2023, Intercom 2024, Harvard Business Review 2020, Freshdesk Benchmarks

All targets are derived conservatively from published benchmarks.

Actual results will depend on cluster coverage and FAQ deployment speed.

By focusing on measurable efficiency and satisfaction metrics, we can validate the ROI of its data-driven support improvements within the first 6 months.

ROI is defined via efficiency and retention improvements, not cost reduction — reinforcing hosting's quality-driven growth strategy.

$ROI \approx (\text{Net Savings from Reduced Workload} + \text{Revenue from Retention Gain}) / \text{Project Cost}$

