ML Platform for Dream11

Rahul Paul

Purpose

- Accelerate in-house ML-driven innovation
- Enhance user experiences
- Secure competitive leadership

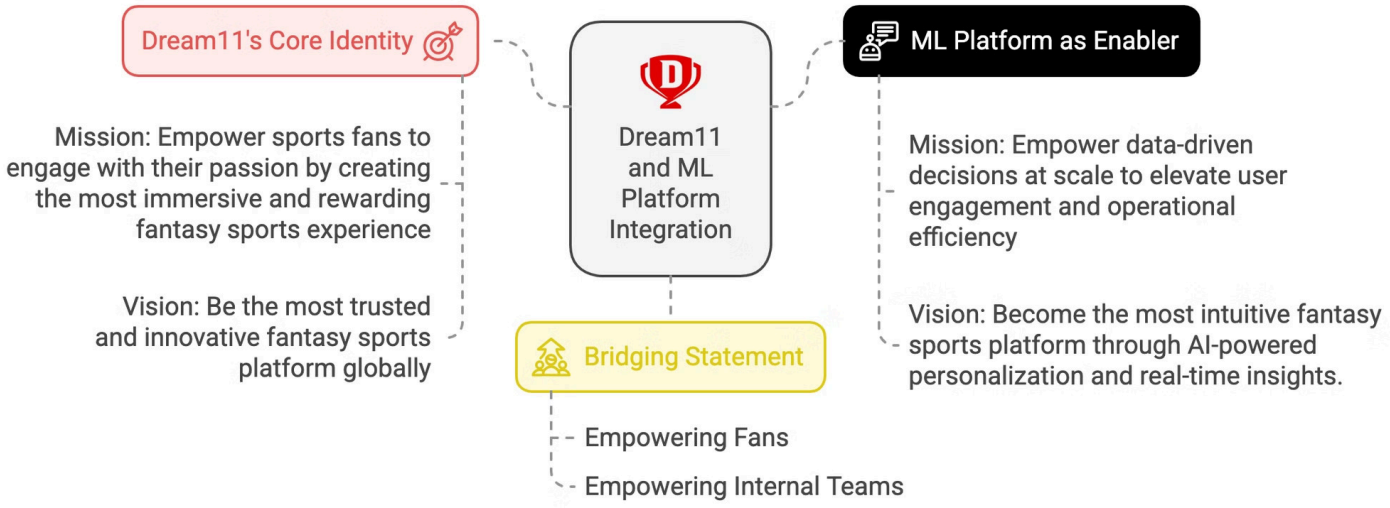
Assumptions

- Availability of historical user/player data
- Existing tech stack can integrate with cloud-based ML tools
- No legal barriers to using AI
- Users will trust and engage with ML-driven features
- May take 12-18 months post launch to justify costs



Why does Dream11 need an ML Platform ?

"Just as Dream11 empowers fans, our ML Platform empowers teams to turn data into delight."



Strategic Importance

- Address scalability during peak events (e.g. IPL)
- Counter rising competition by providing hyper personalized experiences
- Facilitate internal teams to build, train and deploy ML models with ease
- Provide a building block for enabling new use cases like automatic fraud detection, VR experiences etc

Potential Impact

Internal Users

Accelerated Model Deployment

Reduce time-to-production for ML models

Sample Metric: 70% faster iteration cycle for data scientists

Cost Reduction

Automate manual workflows

Reduce reliance on 3P platforms and associated costs

Sample Metric: $2M annual savings from reduced operational overhead

Cross-Team Collaboration

Unified platform for data scientists, developers, and business teams to share insights

Competitive Moats

Proprietary ML models in sports personalizations

Handle 10M+ concurrent users during IPL with real-time predictions

External Users

Hyper-Personalized Experiences

Personalized recommendations, predictive insights, and targeted promotions

Sample Metric: 25% increase in DAU from personalized contest recommendations

Increased User Engagement

Predictive nudges (e.g. "Your star player is trending!") improve stickiness

Sample Metric: 10% improvement in 90-day retention.

Revenue Growth

Dynamic pricing and premium features (e.g., "AI Coach") drive monetization.

Sample Metric: 15% uplift in ARPU

Trust & Fair Play

AI-driven fraud detection reduces collusion, improving platform credibility.

Sample Metric: 40% drop in user complaints about unfair contests.

Key Market Gaps

Lack of Vertical Expertise

Existing tools don't account for fantasy sports nuances (e.g. player substitutions, contest types)

Example: AWS SageMaker can't auto-adjust recommendations for T20 vs. Test cricket dynamics

Fragmented User Experience

Data scientists, developers, and business teams use disjointed tools that slow collaboration

Scalability Limitations

Third-party solutions struggle with traffic spikes (e.g. 10M+ users during IPL finals)

High cost, limited customization

Dream11's Need Gap

- Tailor models to fantasy sports dynamics (e.g., player form, weather)
- Unify cross-functional needs (data scientists + developers + business teams)
- Control over data, infrastructure, and compliance

Comparison with Databricks

Why Build in-house ?

Parameter	Existing Solutions (AWS SageMaker, Google Vertex AI, Third-Party Tools)	In-house platform
Customization	Generic ML frameworks; limited sport-specific optimizations (e.g. player form, weather impact)	Tailored to fantasy sports dynamics (e.g. IPL seasonality, user behavior patterns)
Integration Complexity	Requires effort to integrate with Dream11's existing systems (e.g. payment gateways, user DBs)	Native integration with existing tech stack
Latency and Real-Time Needs	Optimized for batch processing, struggles with <100ms predictions during peak traffic for Dream11 specific scenarios like IPL	Built for real-time inference at scale
Cost Efficiency	High licensing fees + hidden costs (e.g. data egress charges, premium support	Lower total cost of ownership) via cloud-agnostic architecture
Data Privacy & Compliance	Third-party data handling risks (GDPR/PDP compliance dependencies)	Full control over data anonymization and residency (India-specific compliance)
Cross-Team Usability	Separate tools integrated for data scientists, developers, and business teams	Unified platform with role-based dashboards (e.g. marketing analytics, developer health metrics)
Fraud Detection	Generic anomaly detection without fantasy sports context (e.g. collusion patterns in contests)	Proprietary models trained on Dream11's historical fraud data (e.g. multi-account detection)
Time-to-Market	Long setup cycles for sport-specific use cases (e.g. dynamic pricing models)	Once built, enables rapid prototyping (e.g. win probability engine)



Technical and Operational Constraints

Technical Constraints

Integration with existing infrastructure, data pipelines, and security protocols

Time Constraints

- Rapidly evolving technology landscape requires more personalized experiences for users

Cost Constraints

- Engineering build cost
- Operating cost (Data storage + Model Dev + Model Training)
- Deployment and running cost

Assumptions

- Talent availability for ML engineers
- Data accessibility
- Sufficient computing resources for platform development and deployment
- Regulatory approval

User Personas and Jobs to Be Done

Technical Teams

1 Data Scientists

- *Develop and deploy ML models faster*
- *Track performance of deployed models to determine optimizations*

2 Developers

- *Monitor platform health, latency, and performance*
- *Integrate ML models into various Dream11 applications*

Leadership

1 Management

- *Access real-time business insights*
- *Approve budget allocations and evaluate the impact of ML initiatives on key business metrics*

Go-to-Market Teams

1 Marketing Teams

- *Perform user segmentation to derive insights for better targeted marketing campaigns*
- *Campaign performance analysis*

2 Sales Teams

- *Utilize platform data for performance insights and sales forecasting*
- *Track contest performance and pricing strategies*

User Personas Deepdive

Platform Overview and Capabilities

1

Data Ingestion

Secure and efficient data ingestion from various sources, including internal databases, APIs, and external data providers

2

Data Preparation and Feature Engineering

Preprocessing, cleaning, transformation, and feature engineering for optimal model performance

3

Model Development and Training

A wide range of ML algorithms (including AutoML for rapid model prototyping), tools, and libraries for model development and training, including deep learning frameworks

4

Model Deployment and Monitoring

Automated deployment of trained models into production environments (A/B testing + Auto scaled launches) with robust monitoring (Real-time dashboards for system health), versioning, and performance tracking

5

User Interface and Collaboration

An intuitive user interface including Pre-built analytics templates for data scientists, marketing and sales teams to interact with the platform and access insights

6

Integration

Integration with existing Dream11 applications and systems for seamless model utilization and data sharing

System Diagram | Wireframes

Prioritization

Broke down broader features into sub-features. Scored each sub-feature on a scale of 1-5 along the following four parameters:

- Business Impact (BI)
  - 1 - Minimal or indirect impact
  - 5 - Directly impacts revenue/user retention
- User Value (UV)
  - 1 - Nice-to-have for a niche audience
  - 5 - Solves a critical pain point for primary users (data scientists, developers)
- Technical Feasibility (TF)
  - 1 - Requires significant R&D or external expertise
  - 5 - Easy to implement with existing resources
- Time to Market (TTM)
  - 1 - Requires >6 months
  - 5 - Can be delivered in <3 months

Prioritization Deepdive

Product Roadmap: V1, V2, and V3

Feature Prioritization

1

V1: Core Functionality

- Data Ingestion: Real-time Data Streaming, Batch Processing Pipelines, Data Validation & Quality Checks, API Connectors
- Data Preparation & Feature Engineering: Automated Data Cleaning, Feature Transformation
- Model Development & Training: AutoML for Rapid Prototyping, Pre-Built ML Libraries

2

V2: Advanced Features

- Data Preparation & Feature Engineering: Feature Store, Time-Series Feature Engineering
- Model Development & Training: Hyperparameter Tuning, Distributed Training
- Model Deployment & Monitoring: A/B Testing Frameworks, Real-Time Monitoring Dashboards, Auto-Scaling
- User Interface and Collaboration: Data Scientist Workspace, Business User Dashboards

3

V3: Scalability and Optimization

- Data Ingestion: Data Lineage Tracking
- Data Preparation & Feature Engineering: Automated Feature Selection
- Model Development & Training: Deep Learning Frameworks
- Model Deployment & Monitoring: Model Versioning
- User Interface and Collaboration: Collaboration Tools, Role-based access control
- Integration: APIs for Model Consumption, Legacy System Connectors, CI/CD Pipelines

Costing Estimation (v1)

$3.3M (Inhouse-dev)

$7.4M (Third party solutions)

Detailed cost estimation: [**Link**]

Why In-House?

- Long-Term Savings: 50% lower Total Cost of Setup
- Customization: Avoid generic solutions (e.g. sport-specific fraud detection)
- Control: No hidden costs (e.g. data egress fees, premium support)

Key Funding Rationale

- Competitive Moats:
  - Proprietary ML models developed are differentiators
- Scalability:
  - In-house infrastructure handles 10x user spikes (IPL) vs. third-party latency issues at 1/3rd the cost of global cloud regions
- ROI Timeline:
  - Phase 1 ROI: $1M saved via fraud reduction and open-source adoption
  - Phase 2 ROI: 10–15% revenue growth from personalized contests

Plug-and-Play Solutions for Speed

Phase 1

V1: Cloud ML Platforms

Leverage existing cloud-based ML platforms for initial model training and deployment, while focusing on custom features in future versions

Example: Use AWS SageMaker/ Google BigQuery for model training, Open-source data tools (Airflow, Kafka)

Phase 2

V2: Open-Source Libraries and Tools

Integrate open-source libraries and tools for model monitoring, versioning, and collaboration, leveraging existing expertise and reducing development costs

Example: Integrate TensorFlow Extended (TFX) / KubeFlow for pipeline orchestration, Fraud detection system (custom-built), Monitoring tools (Prometheus, Grafana)

Phase 3

V3: In-House Development and Optimization

Prioritize in-house development for scalability, advanced optimization, and integration with Dream11's specific needs

Example: Build custom APIs on Kubernetes for scalability, legacy system integration (Indian IT partners), MLops (CI/CD with Jenkins), AI content tools (fine-tuning existing models)

Anticipating Challenges

Risk	Mitigation Strategy
Talent Attrition	Partner with Indian universities (IITs, IIITs) for internships and pipelines
Data Localization	Use AWS Mumbai region for data residency compliance
Tool Reliability	Hire DevOps experts to manage open-source tools (e.g., Airflow, Feast)
Model bias	Regular fairness audits
Regulatory changes	Engage legal advisors early

Ready to experience the thrill ?

Internal PR Note

Measuring Impact (v1)

1

North Star Metric

% of Data Scientists Actively Using the Platform:

- **Why?** Drives internal adoption and validates the platform's core utility
- **Target:** 70% adoption within 6 months

2

L1

Time-to-first model, Data pipeline uptime, Model Training Speed

3

L2

Avg. model output latency, Beta user retention

4

Guardrail

Cost per model training, System downtime