

Kshitij Goyal

PhD Researcher in Machine Learning

[GitHub](#) | [Personal webpage](#)

Leuven, Belgium

Tel: +32 467 72 88 80

kgoyal40@gmail.com

Experience

PhD Researcher @ DTAI lab, KU Leuven, Belgium

Oct' 2018- present

Topic: Machine Learning with Constraints: Towards Trustworthy AI; *Advisor:* [Prof. Hendrik Blockeel](#)

- **Learning Models that Provably Satisfy Domain Constraints** (*primary goal of the thesis*)
 - Developed a new framework and an optimisation approach to learn models that can **certify domain constraints** (e.g., safety constraints, fairness constraints) for **all possible predictions**.
 - Learning approach uses **combinatorial optimization** for learning the models, and combines it with **gradient descent** to **achieve scalability**.
 - Evaluation on various **regression**, **classification** and **structured prediction** tasks demonstrated that our approach, in contrast to existing approaches like regularization, is able to **certify domain constraints**.
 - Proposed **novel evaluation metrics** to evaluate constraint certification.
- **Automatic Playlist Generation for a Music Streaming Service** (*awarded the **best paper** at BNAIC'22*)
 - Working with **industry partner "Tunify"**, developed and implemented an approach that combines **rule based classification** with **PU learning** to automatically **identify dynamic public playlists**.
 - Proposed a **clustering** based method to **identify new public playlists** from **customer data**, leading to an identification of more than **50 new playlists** previously unidentified by the domain experts.
- **Iteratively Improving Tree Performance by Optimising Subtrees**
 - Proposed and implemented an approach to **improve the performance** of an already learned tree by **optimising sub-trees** iteratively.
 - Demonstrated that the proposed approach improves the performance of CART and lookahead trees to close to **optimal levels**, while being **tractable for deep trees**.
- **Identifying Feature Interaction Constraints to Improve Predictive Performance in Tree Based Models**
 - Conceptualised an approach to use the **feature interactions** from data, identified via a **wrapper approach**, as constraints in the existing XGBoost framework.
 - Interaction Constraints led to an average **improvement of 5%** in the performance for various regression tasks.

Business Analyst - Zynga Games, Bangalore, India

April - Sep' 2017

- Analysed key performance metrics for multiple mobile games to **provide insights** for business strategies in addition to developing an **in-house tool** to perform A/B tests on newly rolled updates.

Business Analyst - Accenture Management Consulting, Bangalore, India

June 2014 - Mar' 2017

- As part of a team, developed **fraud detection techniques** for a reliability management system for an automotive giant to **reduce post-sale expenses**. Proposed approach resulted in a projected **reduction** in warranty spend by **\$249** over the course of 4 years.
- Optimised stock levels at central warehouses across multiple locations for a European telecom giant. **Improved** the total stock value by **9%** by proposing a **rebalancing solution** between different warehouses.

Data Analyst Intern - Media iQ Digital, Bangalore, India

May - July 2013

- **Developed forecasting models** to predict **digital impressions** won by an airline carrier for a given bid.

Education

MSc in Artificial Intelligence, KU Leuven (*graduated magna cum laude*)

Sep' 2017 - Sep' 2018

Master thesis: Proposed and implemented a variant of the classic RankNet approach of ranking documents which personalises the results based on user profiles.

MSc (Integrated) in Mathematics and Scientific Computing, IIT Kanpur, India

June 2009 - May 2014

Master thesis: Analysis of middle censored data under a shifted exponential distribution

Skills

General: Machine Learning · Combinatorial Optimisation · Satisfiability and Logic · Deep Learning · Data Mining · Tree Based Models · Statistics · PU Learning

Programming: Python · SQL · R · Java

Libraries: PyTorch · NumPy · Scikit-learn · Pandas · XGBoost · Altair · z3py · Scoop

Languages: English (Full Professional Proficiency) · Hindi (Native)

Publications

1. SaDe: Learning Models that Provably Satisfy Domain Constraints. ([pdf](#)) *ECML 2022*
K. Goyal, S. Dumancic, H Blockeel
2. Automatic Generation of Product Concepts from Positive Examples. ([pdf](#)) *BNAIC 2022*
K. Goyal, W. Meert, H Blockeel, E. V. Wolputte, K. Vanderstraeten, W. Pijpops, K. Jaspers
3. Feature Interactions in XGBoost. ([pdf](#)) *AIMLAI-ECML 2019*
K. Goyal, S. Dumancic, H Blockeel
4. DeepSaDe: Learning Neural Networks that Guarantee Domain Constraint Satisfaction *(under review)*
K. Goyal, S. Dumancic, H Blockeel

Leadership & Awards

- Teaching assistant for three courses: taught exercise sessions and prepared assignments *2018 - 2022*
- Thesis advisor to 5 students: projects in areas of **personalised search, music streaming, game theory & constrained learning** *2018 - 2022*
- Department representative for the thesis administration for MSc Computer Science at KU Leuven *2019 - 2022*
- Research paper **reviewer** for ECML'19 and ECML'22
- Received the **best paper award** at BNAIC'22
- Awarded **2nd prize** at the KU Leuven Datathon *2017-2018*
- Awarded the prestigious **INSPIRE** scholarship by the government of India for undergraduate studies *2009-2014*